提案50.27。对于Rn上的任意适当凸函数f和任意向量x∈Rn，向量y∈Rn上的下列条件是等价的。

1. y f（x）。
2. 函数hz，yi−f（z）在z=x时达到z的上确界。
3. F（x）+F（y）≤Hx，Yi。（d）f（x）+f（y）=hx，yi。

如果（cl（f））（x）=f（x），那么还有三个条件都与上述条件相等。

（a）x∈f（y）。

（b）函数hx，zi−f（z）在z=y时达到其上确界。

（a）y∈（cl（f））（x）。

以下结果是50.27号提案的推论；见Rockafella[134]（推论23.5.1、23.5.2、23.5.3）。

推论50.28。对于Rn上的任意适当凸函数f，如果f是闭的，则y∈∮f（x）iff x∈∮∮∮（y），对于所有x，y∈rn。

推论50.28表示一种附加性质。

推论50.29。对于Rn上的任何适当凸函数f，如果f在x∈Rn处是可分的，那么（cl（f））（x）=f（x）和（cl（f））（x）=f（x）。

推论50.29表明适当凸函数f的闭包与f一致，其中f是可分的。

推论50.30。对于RN上的任意适当凸函数f，对于RN的任意非空闭凸子集c，对于任意y∈RN，集δ（y c）=（y）由向量x∈RN（如果有的话）组成，其中线性形式z→h7z，yi在c上达到其最大值。

有一个近似次梯度的概念，在优化理论中是有用的；见Bertsekas[19，17]。

定义50.17。设f:rn→r+∞为任意适当的凸函数。对于任意>0，对于任意x∈rn，如果f（x）是有限的，那么f在x上的-次梯度是任意向量u∈rn，这样

，对于所有z∈rn。

见图50.23。f在x处的所有次梯度的集合表示）并称为f在x处的-次微分。

50.4。次微分的附加性质

1)

,

(1

2)

,3/

(0

2,-1)

/

(1

1)

,

(1

(0

2)

,3/

<

z

-

x

,

u

>

+

f

(

x

)

u

=

1

/

2

/

2,-1)

(1

*ε*

f

(

x

)

-

ε

+

<

z

-

x

,

u

>

subgradient

*ε*

-

subgradient

图50.23：设f:r→r−∞，+∞为f（x）=x+1为1定义的分段函数。其铭文为R2中的蓝色阴影区域。线12（x−1）+1（有法向（12−1）是f（x）在（1,1）处的图的支持超平面，而线是与x=1处的-次梯度相关联的超平面，表明u=1∈f（x

集合）可以根据函数hx的共轭定义，由

hx（y）=f（x+y）−f（x），表示所有y∈rn。

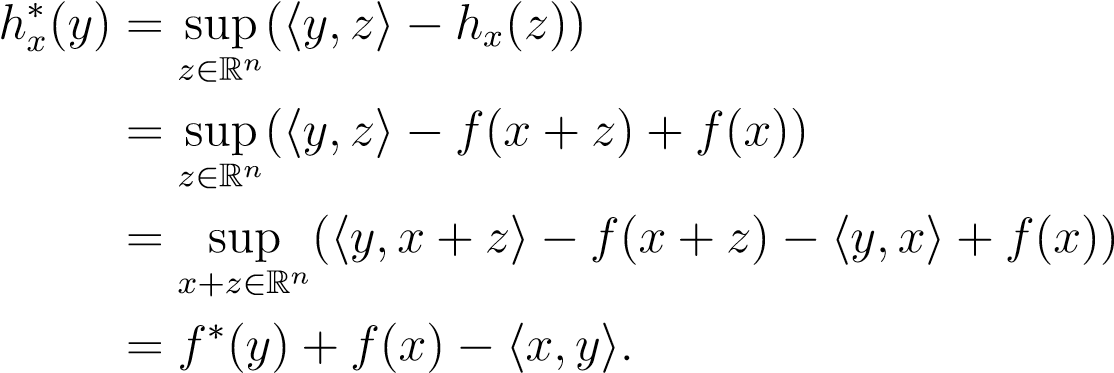
提案50.31。设f:rn→r+∞为任意适当的凸函数。对于任何一个，如果hx由hx（y）=f（x+y）−f（x）给出，对于所有y∈rn，

那么对于所有的y∈rn

和

.

证据。我们有



观察）iff对于每个y∈rn，



敌我识别

.

因为根据定义

h x（u）=sup（h y，ui−hx（y）），y∈rn

我们得出结论

，

如要求。

注：根据Fenchel不等式0和命题50.27（d），消失的向量集是f（x）。

方程表明）是一个闭凸集。

随着变小，集合）减小，我们有

.

但是）不一定减少到δ（y f（x））=i f（x）（y）作为

减小到零。这种差异对应于f0（x；y）和δ（y f（x））=i f（x）（y）之间的差异，这是由于f不一定是封闭的（见命题50.15），如Rockafella[134]中证明的以下结果所示（定理23.6）。

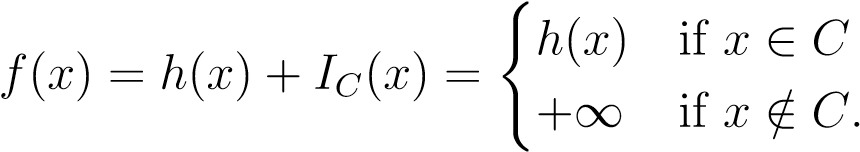
提案50.32。设f是一个闭的、适当的凸函数，设x∈rn使f（x）是有限的。然后

对于所有的y∈rn。

凸函数理论很丰富，我们只给出了一些与优化理论有关的最重要结果的样本。关于凸函数的最小值，还有更多的结果，这些结果因其在优化理论中的应用而特别重要。

# 50.5适当凸函数的最小值

设h为rn上的一个适当凸函数。一般问题是研究RN中非空凸集C上H的最小值，可能由一组不等式和等式约束定义。我们已经观察到，最小化h超过c等于最小化由



因此，首先考虑Rn上的一个适当凸函数f的最小化问题是有意义的。当然，最小化超过rn等于最小化超过dom（f）。

定义50.18。设f为rn上的一个适当凸函数。我们用inf表示

inf f=inf f（x）。

x∈dom（f）

这是函数f对rn的最小值（它可能等于−∞）。对于每个α∈r，我们都有子级集

亚组vα（f）=x∈rn f（x）≤α。

通过命题50.2，我们知道子级集合subsevα（f）是凸的，并且

dom（f）=[亚单位α（f）。阿尔

观察亚组α（f）=∅如果α<inf f。如果inf f>−∞，那么对于α=inf f，亚组α（f）由一组向量组成，其中f达到最小值。

定义50.19。设f为rn上的一个适当凸函数。如果inf f>−∞，则子级集sublvinf f（f）称为f的最小集（此集可能为空）。见图50.24。

(

x,f(x

))

graph of f:

R

->

R

2

y

f (x;y)

≥

0

‘

minimum set of f

x

图50.24：设f为适当的凸函数，其图形是向上的粉红色槽的表面。f的最小集合是r2的浅粉色正方形，它映射到r3中槽的底面。对于最小集合中的任何x，f0（x；y）≥0，由命题50.33证实的事实。

确定最小集是空的还是非空的，或者它是否包含一个点是很重要的。如定理39.11（2）中所述，如果f是严格凸的，那么最小集最多包含一个点。

在任何情况下，我们从命题50.2和50.3知道，f的最小集是凸的，闭的iff f是闭的。

次微分为确定向量x∈rn是否属于f的最小集提供了第一个准则，事实上，子梯度的定义表明x∈rn属于f iff 0∈f（x）的最小集。利用50.15号提案，我们得到以下结果。

提案50.33。设f是Rn上的一个适当凸函数。向量x∈rn属于f iff的最小集合。

0∈f（x）

iff f（x）是有限的，并且

f0（x；y）≥0表示所有y∈rn。

当然，如果f在x上是可微的，那么f（x）=fx，我们得到众所周知的条件fx=0。

有许多方法可以表达命题50.33的条件，f的最小集合甚至可以用共轭函数f来表示。经济衰退方向的概念起着关键作用。

定义50.20。设f:rn→r+∞为任意函数。f的后退方向是任意非零向量u∈rn，这样对于每一个x∈dom（f），函数λ7→f（x+λu）都是非递增的（这意味着对于所有的λ1，λ2∈r，如果λ1<λ2，那么x+λ1u∈dom（f），x+λ2u∈dom（f），和f（x+λ2u）≤f（x+λ1u））。

例50.12。考虑由f（x，y）=2x+y2给出的函数f:r2→r。因为f（x+λu，y+λv）=2（x+λu）+（y+λv）2=2x+y2+2（u+yv）λ+v2λ2，

当v=06时，我们发现λ≥−（u+yv）/v2的上述二次函数增加。如果v=0，则当λ变为+∞时，函数λ7→2x+y2+2uλ减小到−∞；如果u<0，则所有u>0的向量（−u，0）都是衰退方向。见图50.25。

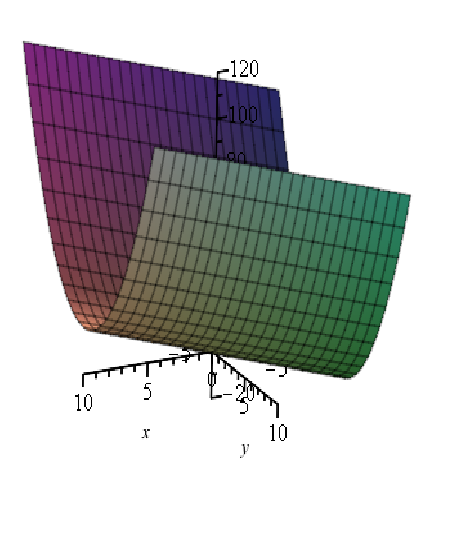
函数f（x，y）=2x+x2+y2没有任何衰退方向，因为f（x+λu，y+λv）=2x+x2+y2+2（u+ux+yv）λ+（u2+v2）λ2，

由于（u，v）=（06，0），我们得到u2+v2>0，因此作为λ的函数，上述二次函数对于λ≥−（u+ux+yv）/（u2+v2）增大。见图50.25。

实际上，上面的例子是典型的。对于任意对称正定n×n矩阵a和任意向量b∈rn，由q（x）=x>ax+b>x给出的二次严格凸函数q没有后退方向。对于任何u∈rn，u=06，我们有

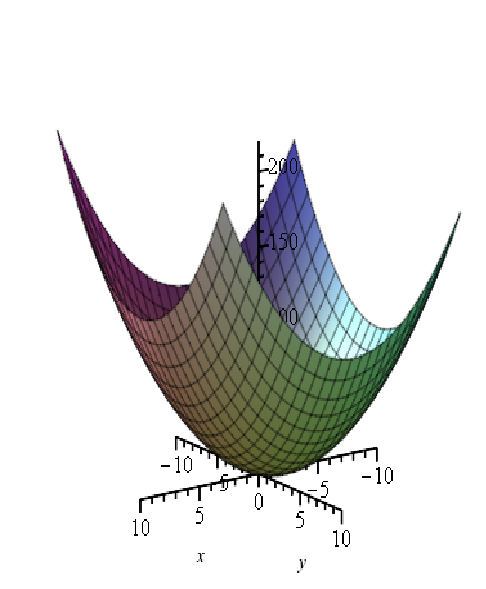
q（x+λu）=（x+λu）>a（x+λu）+b>（x+λu）

=x>ax+b>x+（2x>au+b>u）λ+（u>au）λ2.



f(x,y) = 2x + y

2



f(x,y) = 2x + x + y

2

2

图50.25：示例50.12中讨论的两个函数的图。f（x，y）=2x+y2的曲线沿负x轴“向下”倾斜，反映出（−u，0）是衰退的方向。

由于u=06，a为spd，则u>au>0，且上述二次函数随λ≥−（2x>au+b>u）/（2u>au）而增大。

上述事实给出了凸优化的一个重要技巧。如果f是一个适当的闭凸函数，那么对于任何二次严格凸函数q，函数h=f+q是一个适当的闭严格凸函数，对于一个唯一向量，它具有一个极小值。这个技巧是增强拉格朗日方法的核心，特别是ADMM。令人惊讶的是，严格的证明需要下面的深层定理。

应该注意不要草率地断定，如果凸函数是适当的且是闭的，那么dom（f）和im（f）也是闭的。另外，一个封闭的、适当的凸函数可能不能达到它的最小值。例如，函数f:r→r+∞由

+∞如果x≤0

是一个适当的闭函数和凸函数，但dom（f）=（0，+∞）和im（f）=（0，+∞）。请注意，未达到inf f=0。问题是f的1有一个衰退的方向，如图50.26所示。

以下定理在Rockafella[134]中得到证明（定理27.1）。

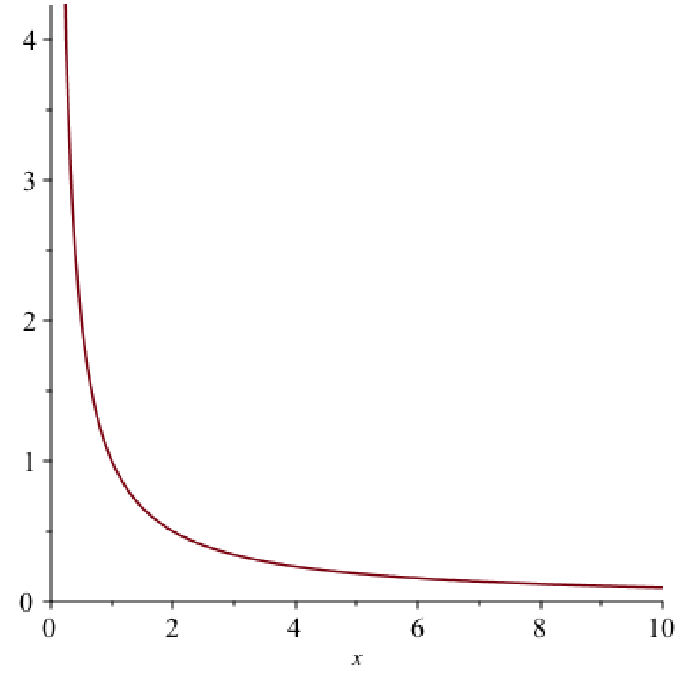


图50.26:x>0的偏函数图。因为1是一个衰退的方向，所以这个函数的图沿着x轴递减。

定理50.34。设f为Rn上的一个适当闭凸函数。以下陈述成立：

1. 我们有inf f=−f（0）。因此f在iff 0∈dom（f）下有界。
2. f的最小集合等于f（0）。因此，得到了f的中值（也就是说有一些x∈rn，这样f（x）=inf f）iff在0处是可分的。当0∈relint（dom（f））时，这种情况尤其适用。此外，当f的每一个衰退方向都是f不变的方向时，0∈relit（dom（f））。
3. 对于f的中值是有限的，但不具绝对性，对于某些y∈rn，f（0）是有限的，（f）0（0；y）=-∞是必要和充分的。
4. f的最小集是一个非空有界集iff 0∈int（dom（f））。这种情况下，如果F没有衰退的方向。
5. f的最小集合由一个唯一的向量x组成，iff在x和

.

1. 对于每一个α∈r，子类α（f）的支持函数是由f+α生成的正齐次凸函数的闭包。如果f在下面有界，则f的最小集合的支持函数是方向导数图y 7→（f）0（0；y）的闭包。

鉴于定理50.34（4）的重要性，我们将此性质表述为以下推论。

推论50.35。设f为RN上的闭正凸函数。那么F的最小集是一个非空有界集，如果F没有衰退的方向。特别地，如果f没有衰退方向，那么f的最小inf f是有限的，并且得到了一些x∈rn。

定理50.13包含以下结果，这对优化程序的设计非常重要。

提案50.36。设f为Rn上的一个适当闭凸函数。将q（x）=x>ax+b>x形式的任意严格凸二次函数q（其中a为对称正定）相加得到的h（x）=f（x）+q（x）所给出的函数h是一个适当的闭严格凸函数，使得inf f是有限的，并且有一个唯一的x∈rn，使得f达到其最小值以x（即f（x）=inf f）表示。

证据。根据定理50.13，有一些仿射形式，由π（x）=c>x+α（带α∈r）给出，这样f（x）≥（x）表示所有x∈rn。然后我们有了

h（x）=f（x）+q（x）≥x>ax+（b>+c>）x+α，所有x∈rn。

由于a是对称正定的，例如50.12，由q（x）=x>ax+（b>+c>）x+α给出的二次函数q没有衰退方向。由于h（x）≥q（x）对于所有x∈rn，我们认为h没有衰退的方向。否则，会有一些非零向量u，这样函数λ7→h（x+λu）对所有x∈dom（h）都是非递增的，所以h（x+λu）≤β对于某些β对于所有λ。但我们证明，对于足够大的λ，函数λ7→q（x+λu）的增加与λ2类似，因此对于足够大的λ，我们将得到q（x+λu）>β，这与H对q的优化相矛盾。根据推论50.35，h具有有限的最小x\_。

如果f和g是适当的凸函数，如果g是严格凸函数，则f+g是适当的函数。对于所有x，y∈rn，对于任何λ，如果0<λ<1，因为f是凸的，g是严格凸的，我们有

F（（1-λ）x+λy）≤（1-λ）f（x）+λf（y）

g（（1-λ）x+λy）<（1-λ）g（x）+λg（y），

因此，我们推断f（（1-λ）x+λy）+g（（1-λ）x+λy）<（（1-λ）（f（x）+g（x））+λ（f（x）+g（x）））），

这表明f+g是严格凸的。然后，由于f+q是严格凸的，它在x处有一个唯一的最小值。

现在我们回到非空凸子集C上的一个适当凸函数h的最小化问题。这是一个很好的描述。

提案50.37。设h为rn上的一个适当凸函数，设c为rn的一个非空凸子集。

1. 对于任何x∈rn，如果有一些y∈h（x），使得−y∈nc（x），即−y在x处对c是正态的，那么h在x处的c上达到其最小值。
2. 如果换衬（dom（h））换衬（c）=6∅，则（1）的倒数保持不变。这意味着，如果h在x上达到c上的最小值，那么有一些y∈h（x），这样−y∈nc（x）。

命题50.37在Rockafella[134]中得到证明（定理27.4）。证明其实很简单。

证据。（1）根据50.33号提案，h在c上的x iff处达到其最小值。

0∈（H+IC）（X）。

根据提案50.22，自

（H+IC）（X）H（X）＋IC（X）、

如果0 h（x）+ic（x），则h在x处的c上达到其最小值。但我们在第50.2节中看到，ic（x）=nc（x），在x处的法锥到c。然后条件0 h（x）+ic（x）表示存在一些y h（x），使得某些z nc（x）的y+z=0，这相当于y∈nc（x）。

（2）根据IC的定义，条件relint（dom（h））relint（c）=6∅是50.22命题的假设

（H+IC）（X）=H（X）+IC（X）、

因此，我们推导出y∈（h+ic）（x），并通过50.33，h在c上达到其在x上的最小值。

注：多面体函数是其上图为多面体的凸函数。很容易看出，50.37（2）号提案也适用于以下情况

1. c为H多面体，换衬（dom（h））c=6∅
2. h为多面体，dom（h）relit（c）=6∅。
3. h和c都是多面体，dom（h）c=6∅。

# 50.6拉格朗日框架的推广

基本上，第49.3节、第49.7节、第49.8节和第49.9节中关于拉格朗日和拉格朗日对偶的所有结果都推广到包含适当和凸目标函数j、适当和凸不等式约束和仿射等式约束的程序。额外的普遍性是不再假定这些函数是

50.6。拉格朗日框架的推广

可微的。这一理论在Rockafella[134]第六部分第28节中对称为普通凸形程序的程序进行了详细讨论。我们没有足够的空间来勾画这个理论，但我们会详细说明一些关键的结果。

我们将处理由一个凸而恰当的目标函数j:rn→r+∞组成的程序，受m≥0不等式约束，其中，i（v）≤0，p≥0仿射等式约束，ψj（v）=0。约束函数i也是凸的和适当的，我们假设relint（dom（j））relint（dom（i））、dom（j）dom（i），i=1，…，m。

这种程序称为普通凸程序。让

u=dom（j）v∈rn i（v）≤0，ψj（v）=0，1≤i≤m，1≤j≤p，

做一套可行的解决方案。我们正在寻找u∈u中使j在u上最小化的元素。

定理49.17的一个广义版本在上述关于j的假设和约束条件i和ψj下成立，但在kkt条件下，涉及梯度的方程必须替换为涉及次微分的下列条件：

，

当i=1，…，m时，λi≥0，j=1，…，p时，m和μj∈r（其中u∈u和j在u处达到其最小值）。

我们问题的拉格朗日L（v，λ，ν）定义如下：

EM＝{X\* RM+P Xi±0, 1±i±m }。

然后

DOM（j）

γ

L（V，λ，礹）=-∞如果（λ，礹）∈/em，V∈dom（j）

+∞如果v/∈dom（j）。

对于固定值，我们还定义了由

，

其有效域为dom（j）（因为我们假设dom（j）dom（i），i=1，…，m）。因此，h（x）=l（x，λ，μ），但h仅是x的函数，因此我们用不同的方式表示它以避免混淆（从技术上讲，l（x，λ，μ）可以取值−∞，但h不取）。

因为j和\_i是适当的凸函数，而ψj是仿射函数，所以函数h是适当的凸函数。

通过将定理28.1、定理28.2和定理28.3放在Rockafella[134]中，可以得到定理49.17的广义版本的证明。为了完整性，我们陈述了这些定理。这是定理28.1。

定理50.38。（定理28.1，Rockafella）让（p）是一个普通的凸规划。设为拉格朗日乘子，使得函数h=j+的中值是有限的，等于j在u上的最佳值。设d为

Rn上h的极小集，设i=i∈1，…，mλi=0。如果d0是d的子集

由向量x组成，使得

|  |  |
| --- | --- |
| \_i（x）≤0 | 就我而言 |
| ⑨i（x）=0ψj（x）=0 | 对于所有i/∈i对于所有j=1，…，p， |

那么，d0是u上（p）的极小值集。

这里是定理28.2。

定理50.39。（定理28.2，Rockafella）让（p）是一个普通的凸规划，让i 1，…，m是非仿射不等式约束的指数的子集。假设（p）的最优值是有限的，并且（p）至少有一个可行解x∈relint（dom（j）），这样

\_i（x）<0表示所有i∈i。

然后存在一些拉格朗日乘数（不一定是唯一的），这样

（a）函数的中值是有限的，等于j在u上的最优值。

定理50.39的假设是约束条件的限定条件，本质上是来自定义49.6的斯莱特条件。

定义50.21。设（p）为普通凸规划，设i 1，…，m为非仿射不等式约束的指数的子集。约束是限定的，是否存在一个可行的解x∈relint（dom（j）），使得

\_i（x）<0表示所有i∈i。

最后，这里是Rockafella[134]的定理28.3。

定理50.40。（定理28.3，Rockafella）让（p）是一个普通的凸规划。如果x∈rn，那么（λ，μ）和x具有

50.6。拉格朗日框架的推广

1. 函数的中值是有限的，等于j在u上的最优值，并且
2. 向量x是（p）（so x∈u）的最优解，iff（x，λ，μ）是（p）的拉格朗日L（x，λ，μ）的鞍点。

此外，如果以下KKT条件成立，则该条件成立：

，且当i=1，…，m时，λi\_i（x）=0。

（2）对于j=1，…p，ψj（x）=0。

.

注意，根据定理50.39，如果（p）的最佳值是有限的，并且约束是合格的，则定理50.40的条件（a）适用于（λ，μ）。因此，我们得到了库恩和塔克定理50.40的以下推论，这是该理论的主要结果之一。它是定理49.17的一个推广版本。

定理50.41。（定理28.3.1，rockafellar）让（p）是一个满足定理50.39假设的普通凸规划，这意味着（p）的最优值是有限的，并且约束是合格的。为了使向量x∈rn是（p）的最优解，存在拉格朗日乘子（λ，祄）是（x，λ，祄）是L（x，λ，祄）的鞍点，这是必要和充分的。等价地，如果且仅当存在拉格朗日乘子时，x是（p）的最优解，它与x一起满足定理50.40中的KKT条件。

定理50.41与（p）的最优解的存在性有关，但它没有说明（p）的最优值。为了建立这样一个结果，我们需要对偶函数的概念。

双功能G的定义如下：

g（λ，μ）=inf l（v，λ，μ）。Ⅴrn

它是一个凹函数（所以−g是凸函数），可以取±∞。请注意，最大化g（相当于最小化−g）会遇到问题，如果对于某些λ，礹，g（λ，礹）=+∞但g（λ，礹）=-∞不会导致问题。乍一看，这似乎违反直觉，但请记住，G是凹的，而不是凸的。凸的是−g，并且−∞和+∞被翻转。

然后，定理49.18（2）的一个更广、更强的版本也成立。通过将推论28.3.1、定理28.4和推论28.4.1放在Rockafella[134]中，可以得到一个证明。为了完整起见，我们陈述了Rockafella[134]的以下结果。

定理50.42。（定理28.4，Rockafella）让（p）是一个具有拉格朗日l（x，λ，μ）的普通凸规划。如果拉格朗日乘子和向量x∈rn具有

1. 函数的中值是有限的，等于j在u上的最优值，并且
2. 向量x是（p）（so x∈u）的最优解，

则鞍值L（x，λ，）为（p）的最佳值J（x）。

一般来说，拉格朗日乘数具有（a）iff的性质。

−∞<inf l（x，λ，）≤suinf l（x，λ，）=inf supl（x，λ，）

xλ，礹xλ，礹

在这种情况下，极值的公共值是（p）的最佳值。特别是，如果x是（p）的最优解，那么supλ，μg（λ，μ）=l（x，λ，μ）=j（x）（零对偶间隙）。

注意定理50.42给出了对偶间隙为零的充分条件（a）和（b）。根据定理50.40，这些条件等价于（x，λ，）是l的鞍点，或等价于KKT条件所具有的事实。

同样，根据定理50.39，如果（p）的最佳值是有限的，并且约束是合格的，则定理50.42的条件（a）适用于（λ，μ）。定理50.42的以下推论成立。

定理50.43。（定理28.4.1，rockafellar）让（p）是一个满足定理50.39假设的普通凸规划，这意味着（p）的最优值是有限的，并且约束是合格的。拉格朗日乘子具有函数中值为有限且等于j在u上的最优值的性质，正是G得到的对偶函数是r的上确界的向量。

定理50.43是定理49.18（2）的一个广义的、更强的版本。定理49.18的第（1）部分要求j和\_i是可微的，因此它不概括。

关于普通凸规划和另一类广义凸规划，可以给出更多的结果。然而，我们并不需要这样的结果来达到我们的目的，特别是讨论ADMM方法。感兴趣的读者可参考Rockafella[134]（第六部分第28和29节）。

# 50.7总结

本章的主要概念和结果如下：

* 扩展实值函数。

50.7。总结

* 铭文（epi（f））。
* 凸凹（扩展实值）函数。
* 有效域（dom（f））。
* 适当和不适当的凸函数。
* 子级集合。
* 下半连续函数。
* 下半连续船体；凸函数的闭合。
* 相对内部（换衬（c））。
* 指示器功能。
* 利普希茨条件。
* 仿射形式，仿射超平面。
* 半个空格。
* 支持超平面。
* A处的法向圆锥。
* 次梯度，次梯度不等式，次微分。
* 闵可夫斯基支持超平面定理。
* 单侧方向导数。
* 支持功能。
* Relu功能。
* -次梯度。
* 凸函数的最小集。
* 衰退的方向。
* 普通凸规划。
* 一套可行的解决方案。
* 拉格朗日
* 鞍点。
* KKT条件。
* 限定约束。
* 二元性差距。

第51章

# 双上升法

本章介绍了目前已知的解决等式约束优化问题的最佳方法之一。实际上，这种方法还可以处理更一般的约束，即凸集的成员关系。它也可以用来解决套索最小化问题。为了更好地理解这种方法，称为乘法交替方向法，简称为ADMM，我们回顾了ADMM的两个前兆：双上升法和乘法方法。

ADMM不是一种新方法。事实上，它是在20世纪70年代发展起来的，由于它非常适合于分布（凸）优化，因此在处理非常大的数据的统计和机器学习问题上，它又重新成为一种非常有效的方法。Boyd、Parikh、Chu、Peleato和Eckstein的优秀论文[28]广泛介绍了ADMM及其变体及其应用。本文实质上是一本关于ADMM主题的书，我们的论述深受其启发。

在本章中，我们考虑在等式约束ax=b下最小化凸函数j（不一定可微）的问题。在第51.1节中，我们讨论了双上升法。它本质上是应用于对偶函数G的梯度下降，但由于G是最大的，梯度下降变成梯度上升。

为了使双上升法的最小化步骤更为稳健，可以使用在拉格朗日中加入惩罚项的技巧。我们得到了增强拉格朗日

，

用λ∈Rm，其中ρ>0称为惩罚参数。我们得到了最小化问题（pρ），

根据au=b最小化，

相当于原始问题。

一千七百一十七

加上惩罚项的好处（根据命题50.36，问题（pρ）在温和条件下有一个唯一的最优解。应用于（pρ）的对偶的对偶上升称为乘数法，并在第51.2节中讨论。

乘法交替方向法，简称ADMM，结合了双重上升的可分解性和乘法方法的优越收敛性。其思想是将函数j分解为两个独立的部分，即j（x，z）=f（x）+g（z），并考虑最小化问题（padmm）。

最小化f（x）+g（z），以ax+bz=c为准，

对于一些p×n矩阵a，一些p×m矩阵b，以及x∈rn，z∈rm，c∈rp，我们也假设f和g是凸的。稍后将添加其他条件。在乘子方法中，我们形成了增广拉格朗日

，

对于一些ρ>0，用λ∈rp。与乘法器方法的主要区别在于，ADMM不是在x和z上共同执行最小化步骤，而是先执行x最小化步骤，然后执行z最小化步骤。因此，x和z以交替或顺序的方式更新，这解释了术语交替方向。由于拉格朗日是增强的，因此A和B上的一些温和条件意味着这些最小化步骤将被保证终止。ADMM见第51.3节。

在第51.4节中，我们在以下假设下证明了ADMM的收敛性：（1）函数f:r→r+∞和g:r→r+∞是适当的闭凸函数（见第50.1节），这样relit（dom（f））relit（dom（g））=6∅。

1. n×n矩阵a>a是可逆的，m×m矩阵b>b是可逆的。等价地，p×n矩阵a具有秩n，p×m矩阵具有秩m。
2. 未经认可的拉格朗日l0（x，z，λ）=f（x）+g（z）+λ>（ax+bz−c）具有鞍点，这意味着存在x，z，λ（不一定是唯一的），因此

l0（x，z，λ）≤l0（x，z，λ）≤l0（x，z，λ）

对于所有x，z，λ。

根据定理50.40，假设（3）等价于KKT方程被一些三重方程（x，z，λ）满足的事实，即

ax+bz−c=0（）

和

0∈f（x）＋g（z）＋a>λ+b>λ，（†）

51.1。双上升

假设（3）也等价于定理50.40的条件（a）和（b）。特别是，我们的程序有一个最佳解决方案（x，z）。根据定理50.42，λ是对偶函数g（λ）=infx，z l0（x，z，λ）的最大值，强对偶性保持，即g（λ）=f（x）+g（z）（对偶间隙为零）。

在定理51.1的证明之后，我们将证明假设（2）实际上是由假设（3）所隐含的。这使得我们能够证明一个比Boyd等人所证明的收敛结果更强的收敛结果。【28】（在完全相同的假设（1）和（3）下）。特别地，我们证明了所有序列（xk）、（zk）和（λk）都收敛到最优解（x，）和λe。我们证明的核心在于Boyd等人。[28]但是有新的步骤，因为我们有更强有力的假设（2）。

在第51.5节中，我们讨论了停止标准。

在第51.6节中，我们介绍了ADMM的一些应用，特别是在RN中，在闭凸集C上最小化适当的闭凸函数f，以及二次规划。第二个例子提供了解决二次问题的最佳方法之一，特别是第54章讨论的SVM问题。

第51.7节给出了ADMM在“1-范数问题”中的应用，特别是在机器学习中起重要作用的lasso正则化。

## 51.1双升

我们的目标是解决最小化问题，问题（P）。

根据au=b，将j（u）最小化，

具有仿射等式约束（具有m×n矩阵和b∈rm）。问题（p）的拉格朗日L（u，λ）由下式给出：

l（u，λ）=j（u）+λ>（au-b）。

用λ∈Rm。从命题49.19出发，对偶函数g（λ）=infu∈rn l（u，λ）由下式给出：

（

−b>λ−j（−a>λ）如果−a>λ∈dom（j），

G（λ）=

−∞否则，

对于所有的λ∈rm，其中j是j的共轭。回想一下，根据定义49.11，在rn的子集u上定义的函数f:u→r的共轭f是由定义的部分函数f：rn→r。

.

如果定理49.18（1）的条件成立，在我们的例子中，这意味着对于每个λ∈Rm，都有一个唯一的uλ∈Rn，这样

g（λ）=l（uλ，λ）=infn l（u，λ），u∈r

函数λ7→uλ是连续的，那么g是可微的。此外，我们有

gλ=auλ−b，

对于双问题的任何解，μ=λ

最大化g（λ），服从于λ∈Rm，

向量u=u祄是原始问题（p）的一个解决方案。此外，j（u）=g（λ），即二元间隙为零。

双上升法本质上是应用于双功能G的梯度下降，但由于G最大化，梯度下降成为梯度上升。另外，我们不再担心上述问题的极小问题inf+一些极小值，namelyu∈rn l（u，λ）有一个唯一的解，因此我们用

U

u+=argminl（u，λ）。

U

给定初始双变量λ0，双上升法包括以下两个步骤：

UK+1=阿格明（u，λk）

，

其中αk>0是一个步长。第一步用于计算“新梯度”（实际上，如果Minimizer UK+1是唯一的，那么gλk=a uk+1−b），第二步是双变量更新。

例51.1。让我们来看一个非常简单的例子，即应用于我们在第41.1节中第一次遇到的问题的梯度上升法，即最小化j（x，y）=（1/2）（x2+y2），服从于2x−y=5。拉格朗日是

.

见图51.1。

拉格朗日对偶的方法是首先计算g（λ）=inf l（x，y，λ）。

（x，y）∈r2

51.1。双上升

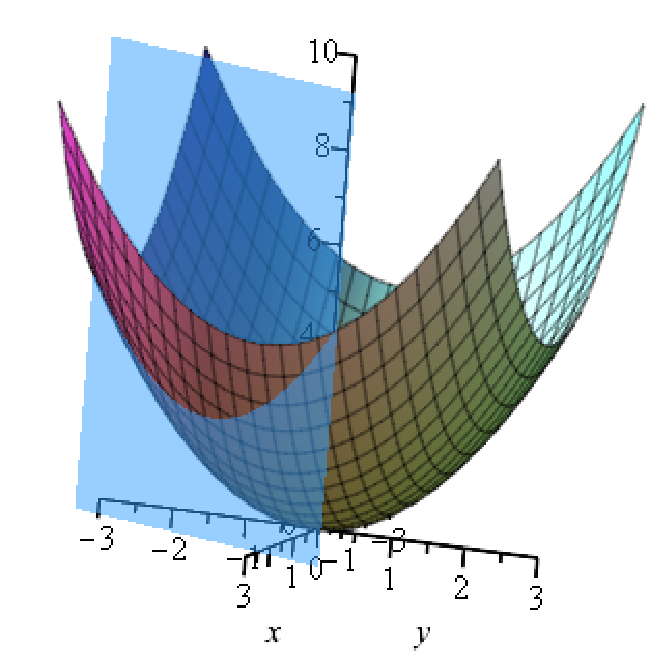


图51.1:j（x，y）=（1/2）（x2+y2）为抛物面，2x−y=5为透明蓝色。实施例51.1的解是交叉曲线的顶点，即点（2，

自从

，

我们发现j（x，y）是由正定矩阵决定的二次函数。

，因此要计算g（λ），我们必须设置lx，y=0。通过计算=0和

=0，我们发现x=−2λ和y=λ。然后g（λ）=−5/2λ2−5λ，我们必须计算g（λ）相对于λ∈r的最大值。这意味着计算g0（λ）=0并获得（x，y，λ）=（−2λ，λ，λ）=（2、−1、−1）的解。

对偶同意法不是直接用（x，y）来求解λ，而是从λ的数值估计开始，即λ0，形成“数值”拉格朗日。

.

利用这个数值λ0，我们将l（x，y，λ0）与（x，y）的关系最小化。该计算将与上面形成g（λ）所用的计算相同，因此，我们得到迭代步骤（x1，y1）=（-2λ0，λ0）。因此，如果我们用λk替换λ0，我们就得到了双上升法的第一步，即

.

双上升法的第二步通过计算改进了λ的数值估计。

.

（回想一下，在我们最初的问题中，约束是2x−y=5或5，所以

通过简化上述方程，我们发现

λk+1=（1−β）λk−β，β=5αk。

用前面方程中的λk进行反代换，结果表明：

λk+1=（1−β）k+1λ0+（1−β）k+1−1。

如果0<β≤1，前一行意味着λk+1收敛于λ=−1，这与原拉格朗日对偶法提供的答案一致。观察β=1或，双升法一步结束。

选择适当的αk，我们得到了g（λk+1）>g（λk），该方法取得了进展。例如，在某些假设下，J是严格凸的，并且αk的某些条件下，可以证明双上升收敛到最优解（对于原解和对偶解）。然而，双重上升的主要缺陷是极小化步骤可能出现分歧。例如，发生这种情况的是j是它的一个分量的非零仿射函数。补救办法是给拉格朗日加上一个惩罚条款。

在积极方面，如果函数j是可分离的，则双重上升法将导致分散算法。假设u可以拆分为，使用ui∈rni和

，那

n

j（u）=xji（ui）

i＝1

A被分成n个块ai（ai是m×ni矩阵），作为a=[a1····an]，所以。那么拉格朗日可以写成

具有

.

由此可知，关于原变量u的l（u，λ）的最小化可以分解为n个单独的最小化问题，这些问题可以并行解决。然后算法执行n个更新

=阿格米利（Ui，λk）

用户界面

平行，然后是台阶

λk+1=λk+αk（auk+1−b）。

## 51.2增广拉格朗日和乘数法

为了使双上升法的最小化步骤更为稳健，可以使用在拉格朗日中加入惩罚项的技巧。

定义51.1.考虑到优化问题（p），

根据au=b，将j（u）最小化，

增广拉格朗日由下式给出

，

用λ∈Rm，其中ρ>0称为惩罚参数。

增广拉格朗日Lρ（u，λ）可以看作是极小问题（pρ）的普通拉格朗日。

根据au=b最小化。

上述问题相当于程序（p），因为对于（pρ）的任何可行解，我们必须有au-b=0。

加上惩罚项的好处（即根据命题50.36，问题（pρ）在a的温和条件下具有唯一的最优解。

对（pρ）的对偶应用的对偶上升产生乘数方法，该方法由以下步骤组成，给定一些初始λ0：

=argminlρ（u，λk）λk+1=λk+ρ（auk+1−b）。

观察第二步使用参数ρ。其原因是，选择αk=ρ可以保证（uk+1，λk+1）满足方程。

juk+1+a>λk+1=0，

这意味着（UK+1，λk+1）是双重可行的；见Boyd、Parikh、Chu、Peleato和Eckstein[28]第2.3节。

例51.2。考虑最小化问题

最小化Y2+2X，以2X−Y=0为准。

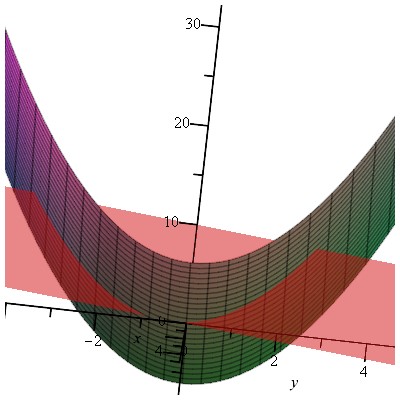
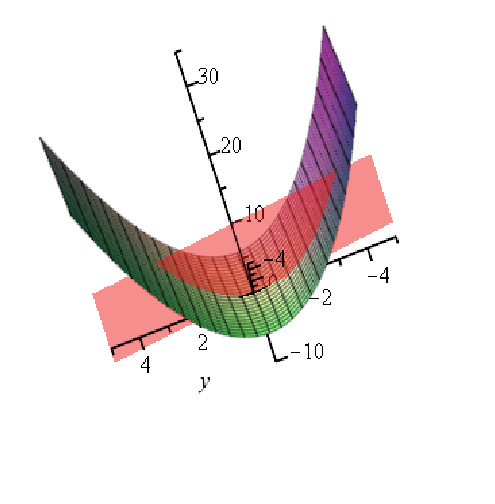


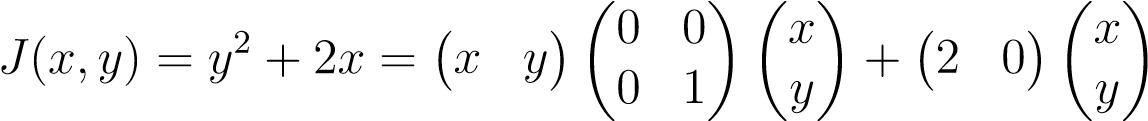
图51.2:Y2+2X图形与透明红色平面相交的两个视图

2x−y=0.例51.2的解是相交曲线的顶点，即点

1 1

见图51.2。

二次函数



是凸的，但不是严格凸的。由于y=2x，问题相当于最小化y2+2x=4x 2+2x，其最小值为x=−1/4（因为设置函数x→7 4x2+2的导数得到8x+2=0）。因此，我们的问题的唯一最小值为（x=−1/4，y=−1/2）。我们问题的最大值是l（x，y，λ）=y2+2x+λ（2x-y）。

如果我们采用双上升法，保持λ常数的l（x，y，λ）对x和y的最小化会产生方程。

2+2λ=0

2y-λ=0，

通过将L的梯度（相对于x和y）设置为零获得。如果λ=6−1，则问题没有解决方案。实际上，如果λ=6−1，将l（x，y，λ）=y2+2x+λ（2x−y）与x和y的关系最小化，则得出−∞。

增强拉格朗日是

，

矩阵形式是

.

上述矩阵的迹线为1+0，行列式为

，

因为ρ>0。因此，上述矩阵是对称正定的。将Lρ（x，y，λ）对x和y最小化，将Lρ（x，y，λ）（对x和y）的梯度设为零，得到方程：

解决办法是

.

因此，乘数法的步骤是

，

第二步简化为

λk+1=−1.

因此，我们发现对于任何初始值λ0，该方法在两步后收敛，我们得到

.

乘法的方法也收敛于非凸函数j，如下一个例子所示。

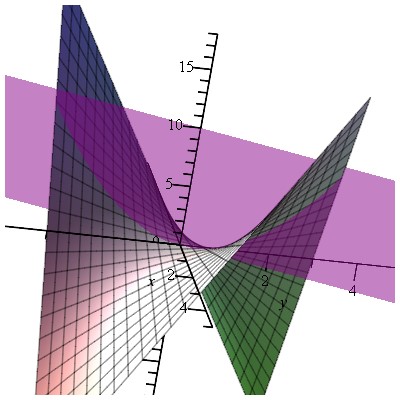
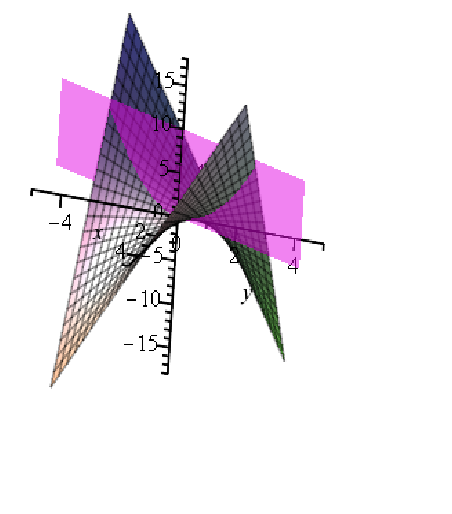
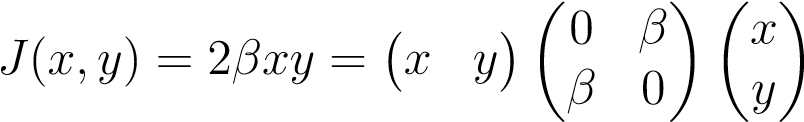


图51.3:2xy（β=1）鞍形图的两个视图与透明品红色平面2x−y=0相交。例51.3的解是相交曲线的顶点，即点（0,0,0）。

例51.3。考虑最小化问题

最小化2βxy，以2x−y=0为准，

β>0。见图51.3。二次函数



不是凸的，因为上面的矩阵不是正半定的（矩阵的特征值是−β和+β）。增强拉格朗日是

，

矩阵形式是

.

上述矩阵的迹线为20，行列式为

ρ2−（β−ρ）2=β（2ρ−β）。

如果ρ>β/2，这个行列式是正的，在这种情况下，矩阵是对称正定的。将Lρ（x，y，λ）对x和y最小化，将Lρ（x，y，λ）（对x和y）的梯度设为零，得到方程：

2ρx+（β-ρ）y+λ=0

2（β-ρ）x+ρy-λ=0。

因为我们假设ρ>β/2，所以解是

.

因此，乘数法的步骤是

，

第二步简化为

，

也就是说，

-

如果我们选取ρ>β>0，这意味着ρ>β/2，那么

，

该方法收敛于任何初值λ0的解。

x=0，y=0，λ=0.

实际上，由于约束2X−Y=0成立，2βx y=4βx2，并且函数x 7→4βx2的最小值在x=0时实现（因为β>0）。

作为练习，读者应该验证双上升（αk=ρ）产生的方程

，

因此，该方法有分歧，除了λ0=0，这是最优解。

乘法器的方法在比双提升更普遍的条件下收敛。然而，惩罚项的加成具有负效应，即J可分，拉格朗日Lρ也不可分。因此，乘法器的基本方法不能用于分解，也不可并行。下一种方法处理可分离性问题。

## 51.3 ADMM：乘法器的交替方向法

乘法交替方向法，简称ADMM，结合了双重上升的可分解性和乘法方法的优越收敛性。它可以被看作是乘法器方法的近似值，但一般来说它更优越。

其思想是将函数j分解为两个独立的部分，即j（x，z）=f（x）+g（z），并考虑最小化问题（padmm）。

最小化f（x）+g（z），以ax+bz=c为准，

对于一些p×n矩阵a，一些p×m矩阵b，以及x∈rn，z∈rm，c∈rp，我们也假设f和g是凸的。稍后将添加其他条件。在乘子方法中，我们形成了增广拉格朗日

-至-

对于一些ρ>0，用λ∈rp。

给定一些初始值（z0，λ0），admm方法包括以下迭代步骤：

xk+1=argminlρ（x，zk，λk）

X

zk+1=argminlρ（xk+1，z，λk）

.

51.3。ADMM：乘数交替方向法

而不是在x和z上共同执行最小化步骤，就像步骤中的乘数方法一样

（xk+1，zk+1）=argminlρ（x，z，λk），

x，z

ADMM首先执行X最小化步骤，然后执行Z最小化步骤。因此，x和z以交替或顺序的方式更新，这解释了术语交替方向。

ADMM中的算法状态为（zk，λk），从这个意义上说（zk+1，λk+1）是（zk，λk）的函数。变量xk+1是一个辅助变量，用于从（zk，λk）计算zk+1。X和Z的作用并不完全对称，因为X的更新是在λ的更新之前完成的。通过切换X和Z，F和G，A和B，我们得到了一个ADMM的变体，其中X-update步骤和Z-update步骤的顺序是相反的。

例51.4。让我们重新考虑一下示例51.2中的问题，用admm来解决它。我们把问题表述为

最小化2X+Z2，以2X−Z=0为准，

f（x）=2x，g（z）=z2。增广拉格朗日由下式给出

.

ADMM步骤如下。X-update是

xk+1=argmin，

X

由于这是x中的二次函数，当上述函数的导数（相对于x）为零时，即

（1）

z-update是

，

对于x阶跃，当上述函数的导数（相对于z）为零时，即达到最小值。

（2）

|  |  |
| --- | --- |
| λ-更新为 |  |
| λk+1=λk+ρ（2xk+1−zk+1）。 | （3） |

将（1）的右侧替换为（2）中的xk+1，得出

（4）

利用（2），我们得到

，（5）

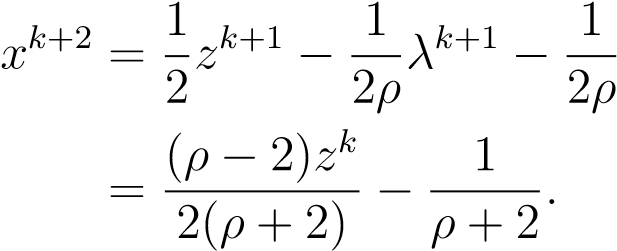
然后用我们得到的

（6）

将（1）的右侧替换为（6）中的xk+1，我们得到

（7）

式（7）表明zk确定了λk+1，式（1）表明k+2，以及式（4）表明zk也确定了xk+2。特别是，我们发现



这样就足以找到该层序的界限（zk）。因为我们已经从例子中知道了

51.2该限值为−1/2，使用（4），我们写下

.

通过归纳，我们推断出

，

由于ρ>0，我们得到ρ/（ρ+2）<1，所以序列的极限（zk+1）确实是−1/2，因此（λk+1）的极限是−1，xk+2的极限是−1/4。

为了使ADMM实用，X-最小化步骤和Z-最小化步骤必须能够有效地实现。

根据标度双变量礹=（1/ρ）λ，编写ADMM更新通常很方便。如果我们将残差定义为

R=ax+bz−c，

然后我们有了

λ>r+（ρ/2）k r k22=（ρ/2）kr+（1/ρ）λk22−（1/（2ρ））kλk22=（ρ/2）kr+µk22−（ρ/2）kµk22。

按比例缩放的ADMM形式包括以下步骤：

xk+1=argmin

X

祄k+1=祄k+axk+1+bzk+1−c。

如果我们将步骤k的剩余Rk定义为

Rk=axk+bzk−c=祄k−祄k−1=（1/ρ）（λk−λk−1），

然后我们看到了

.

按比例排列的公式通常比未按比例排列的公式短。

我们现在讨论ADMM的收敛性。

## 51.4行政管理的衔接

让我们重复一下admm的步骤：给定一些初始值（z0，λ0），请执行以下操作：

xk+1=argminlρ（x，zk，λk）（x-更新）

X

zk+1=argminlρ（xk+1，z，λk）（z-更新）

（λ-更新）

可在以下三个假设下证明ADMM的收敛性：

1. 函数f:r→r+∞和g:r→r+∞是适当的闭凸函数（见第50.1节），这样relit（dom（f））relit（dom（g））=6∅。
2. n×n矩阵a>a是可逆的，m×m矩阵b>b是可逆的。等价地，p×n矩阵a具有秩n，p×m矩阵具有秩m。
3. 未经认可的拉格朗日l0（x，z，λ）=f（x）+g（z）+λ>（ax+bz−c）具有鞍点，这意味着存在x，z，λ（不一定是唯一的），因此

l0（x，z，λ）≤l0（x，z，λ）≤l0（x，z，λ）

对于所有x，z，λ。

回想一下，增广拉格朗日由

.

对于z（和λ）固定，我们有

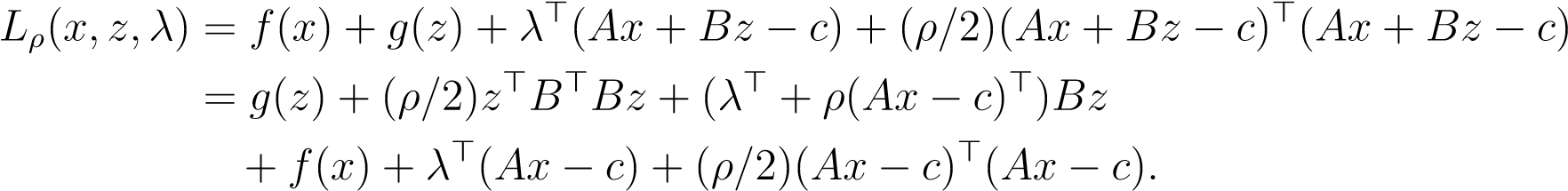
Lρ（x，z，λ）=f（x）+g（z）+λ>（ax+bz−c）+（ρ/2）（ax+bz−c）>（ax+bz−c）

=f（x）+（ρ/2）x>a>ax+（λ>+ρ（bz−c）>）ax

+G（Z）+λ>（BZ−C）+（ρ/2）（BZ−C）>（BZ−C）。

假设（1）和（2）保持不变。由于a>a是可逆的，那么它是对称正定的，并且根据命题50.36，x-最小化步骤有一个唯一的解（最小化问题用一个唯一的最小化器成功）。

同样，对于x（和λ）固定，我们有



由于b>b是可逆的，那么它是对称正定的，并且根据命题50.36，z-最小化步骤有一个唯一的解（最小化问题以一个唯一的最小化器成功）。

根据定理50.40，假设（3）等价于KKT方程被一些三重方程（x，z，λ）满足的事实，即

ax+bz−c=0（）

和

0∈f（x）＋g（z）＋a>λ+b>λ，（†）

假设（3）也等价于定理50.40的条件（a）和（b）。特别是，我们的程序有一个最佳解决方案（x，z）。根据定理50.42，λ是对偶函数g（λ）=infx，z l0（x，z，λ）的最大值，强对偶性保持，即g（λ）=f（x）+g（z）（对偶间隙为零）。

在定理51.1的证明之后，我们将看到假设（2）实际上是由假设（3）所隐含的。这使得我们能够证明一个比Boyd等人所证明的收敛结果更强的收敛结果。[28]在完全相同的假设（1）和（3）下。

设p为凸集上f+g的最小值（x，z）rm+p ax+bz−c=0，设（pk）为pk=f（xk）+g（zk）给出的序列，并回想Rk=axk+bzk−c。

我们的主要目标是证明以下结果。

定理51.1。假设以下假设成立：

1. 函数f:r→r+∞和g:r→r+∞是适当的闭凸函数（见第50.1节），这样relit（dom（f））relit（dom（g））=6∅。
2. n×n矩阵a>a是可逆的，m×m矩阵b>b是可逆的。同样地，p×n矩阵a有秩n，p×m矩阵有秩m（这个假设实际上是多余的，因为它是由假设（3）所隐含的）。
3. 未经认可的拉格朗日l0（x，z，λ）=f（x）+g（z）+λ>（ax+bz−c）具有鞍点，这意味着存在x，z，λ（不一定是唯一的），因此

l0（x，z，λ）≤l0（x，z，λ）≤l0（x，z，λ）

对于所有x，z，λ。

然后，对于任何初始值（z0，λ0），以下属性保持不变：

1. 序列（Rk）收敛到0（剩余收敛）。
2. 序列（pk）收敛到p（目标收敛）。
3. 序列（xk）和（zk）收敛到问题（padmm）的最优解（x，e ze），序列（λk）收敛到对偶问题（初等和对偶变量收敛）的最优解λe。

证据。证据的核心在于Boyd等人。[28]但是有新的步骤，因为我们有更强有力的假设（2），产生更强有力的结果（3）。

证据包括几个步骤。不能直接证明序列（xk）、（zk）和（λk）收敛，因此首先证明序列（rk+1）收敛到零，并且序列（axk+1）和（bzk+1）也收敛。

第1步。证明下面的不等式（a1）。

考虑reals（v k）的序列

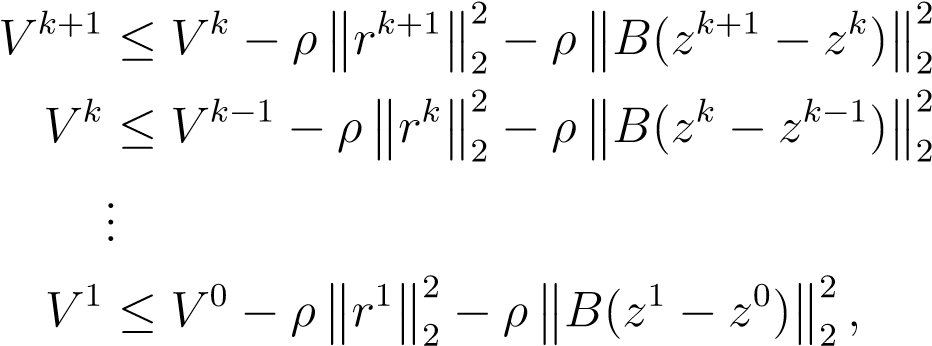
.

可以看出，v k满足以下不等式：

（A1）

这相当困难。因为Boyd等人给出了完整的证据。[28]之后，我们将只提供一些关键步骤。

不等式（a1）表明序列（v k）是不递增的。如果我们为k，k−1，…，0写下这些不等式，我们有



通过把这些不等式相加，我们得到

，

这意味着

，（b）

因为v k+1≤v 0。

第2步。证明了序列（Rk）收敛于0，并且序列（axk+1）和（bzk+1）也收敛。

不等式（b）意味着级数绝对收敛。

特别是，序列（Rk）收敛到0。

系列的第n个部分和）是

，

由于序列）收敛，我们推断序列（bzk+1）收敛。由于axk+1+bzk+1−c=rk+1，（rk+1）和（bzk+1）的收敛意味着序列（axk+1）也收敛。

第3步。证明了序列（xk+1）和（zk+1）收敛。根据假设（2），矩阵a>a和b>b是可逆的，因此将每个向量a xk+1乘以（a>a）−1a>，如果序列（axk+1）收敛到u，那么序列（xk+1）收敛到（a>a）−1a>u。同样，如果序列（b zk+1）收敛到v，那么序列（zk+1）收敛到（b>b）−1B>

第4步。证明序列（λk）收敛。

回想一下

λk+1=λk+ρrk+1.

接下来是归纳法

λk+p=λk+ρ（Rk+1+···+ρk+p），p≥2.

结果，我们得到

.

由于级数收敛，部分和形成一个柯西序列，这立即意味着对于任何>0，我们都可以找到n>0，这样

对于所有k，p+k≥n，

因此序列（λk）也是柯西序列，因此它收敛。

第5步。证明了序列（pk）收敛于p。

为此，我们还需要两个不平等。遵循Boyd等人[28]我们需要证明

pk+1−p≤−（λk+1）>Rk+1−ρ（b（zk+1−zk））>（−Rk+1+b（zk+1−z））（a2）

和

P−pk+1≤（λ）>Rk+1.（A3）

因为我们证明了序列（Rk）和b（zk+1−zk）收敛于0，并且序列（λk+1）收敛于

（λk+1）>Rk+1+ρ（b（zk+1−zk））>（−Rk+1+b（zk+1−z））≤p−pk+1≤（λ）>Rk+1，我们推断在极限处，pk+1收敛于p。

第6步。证明（A3）。

因为（x，y，λ）是鞍点，我们有

l0（x，z，λ）≤l0（xk+1，zk+1，λ）。

由于ax+bz=c，我们得到l0（x，z，λ）=p，由于pk+1=f（xk+1）+g（zk+1），我们得到

l0（xk+1，zk+1，λ）=pk+1+（λ）>rk+1，

因此我们得到p≤pk+1+（λ）>Rk+1，

产生（a3）。

第7步。证明（A2）。

根据50.33号提案，zk+1最小化lρ（xk+1，z，λk）iff

0∈g（zk+1）+b>λk+ρb>（axk+1+bzk+1−c）

=g（zk+1）+b>λk+ρb>Rk+1

=g（zk+1）+b>λk+1，

因为Rk+1=axk+1+bzk+1−c和λk+1=λk+ρ（axk+1+bzk+1−c）。

总之，我们有

0 g（zk+1）+b>λk+1，（†1）

这表明zk+1最小化了函数

z 7→g（z）+（λk+1）>bz。

因此，我们

g（zk+1）+（λk+1）>bzk+1≤g（z）＋（λk+1）>bz（b1）

同样，xk+1最小化lρ（x，zk，λk）iff

0∈f（xk+1）+a>λk+ρa>（axk+1+bzk−c）

=f（xk+1）+a>（λk+ρrk+1+ρb（zk−zk+1））。

=f（xk+1）+a>λk+1+ρa>b（zk−zk+1）

因为Rk+1−Bzk+1=axk+1−c和λk+1=λk+ρ（axk+1+Bzk+1−c）=λk+ρRk+1。

同样，上述推导表明

0∈f（xk+1）+a>（λk+1−ρb（zk+1−zk）），（†2）

这表明XK+1最小化了功能

x 7→f（x）+（λk+1−ρb（zk+1−zk））>ax。

因此，我们得到f（xk+1）+（λk+1−ρb（zk+1−zk））>axk+1≤f（x））+（λk+1−ρb（zk+1−zk））>ax。（B2）

将不等式（b1）和（b2）相加，利用方程ax+bz=c，重新排列，得到不等式（a2）。

第8步。证明（xk）、（zk）和（λk）收敛于最优解。回想一下，（Rk）收敛到0，（Xk）、（zk）和（λk）收敛到极限x、z和

e e

λe.由于Rk=axk+bzk−c，在极限内，我们有

|  |  |
| --- | --- |
| 使用（†1），在极限内，我们获得 |  |
| 0 g（z）+b>λ.e  由于（b（zk+1−zk））收敛到0，使用（†2），在极限内，我们得到 | （2） |
| 0∈f（x）+a>λ.e  从（2）和（3），我们得到 | （3） |
| 0∈f（x）+g（z）+a>λe+b>λ.e e | （4） |

ax+bz−c=0.（1）e

但（）和（4）正是KKT方程，根据定理50.40，我们得出x、z和λ是最优解。EE

第9步。证明（A1）。这是证据中最冗长的一步。我们首先将（a2）和（a3）相加，然后执行相当多的操作或重写操作。完整的推导可在Boyd等人〔28〕。

评论：

1. 根据定理50.41，我们可以用稍微强一些的假设来代替假设（3），即程序的最佳值是有限的，并且约束是合格的。由于约束是仿射的，这意味着在relit（dom（f））relit（dom（g））中存在一些可行的解决方案。这些假设比假设（3）更实际。
2. 实际上，假设（3）意味着假设（2）。实际上，我们从定理50.40知道鞍点的存在意味着我们的程序有一个有限的最优解。但是，如果a>a或b>b不可逆，那么程序（p）可能没有有限的最优解，如下反例所示。

例51.5。让

f（x，y）=x，g（z）=0，y−z=0。

然后

Lρ（x，y，z，λ）=x+λ（y−z）+（ρ/2）（y−z）2，

但用Z保持不变的超（x，y）最小化得到−∞，这意味着上述程序没有有限的最优解。见图51.4。

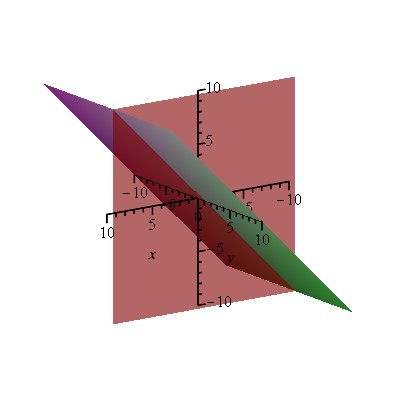
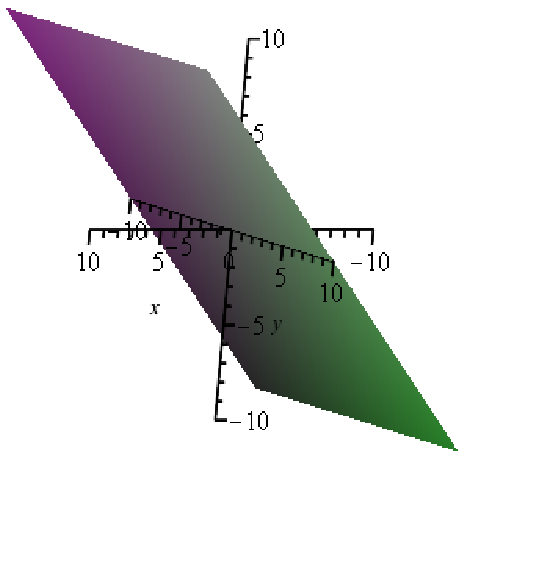
问题是

，

但是

不可逆。

1. 证明（a1）、（a2）、（a3）和（rk）到0和（pk）到p的收敛性不需要假设（2）。使用独创不等式（a1）（和（b））的证明是Boyd等人给出的证明。〔28〕。我们还能够证明（λk）、（axk）和（bzk）在没有假设（2）的情况下收敛，但为了证明（xk）、（yk）和（λk）收敛到最优解，我们必须使用假设（2）。



f(x,y) = x intersected with y=z,

z fixed.

graph of f(x,y) = x

图51.4：示例51.5的图形表示。这是固定Z时X最小化步骤的图示。由于这两个平面的交叉点是一条无界线，我们“看到”最小化x的结果是−∞。

1. Bertsekas在[17]第2.2和5.4节中讨论了ADMM。他对ADMM的表述略有不同，即

最小化f（x）+g（z），以ax=z为准。

Bertsekas在假设dom（f）是紧凑的或a>a是可逆的，并且存在鞍点的前提下陈述了该版本admm的收敛结果；见命题5.4.1。证据见Bertsekas[20]第3.4节，提案4.2。似乎证明利用了梯度，所以不清楚它适用于更一般的情况，其中f和g是不可微的。

1. 在加贝[47]中讨论了ADMM的版本（第4节和第5节）。它们比这里讨论的版本更通用。给出了一些收敛性的证明，但由于加贝的框架更为一般，不清楚它们是否适用于我们的环境。同时，这些证据也依赖于狮和麦赛的早期结果，这使得比较变得困难。

（5）假设（2）并不意味着系统ax+bz=c有任何解决方案。例如，如果

，

系统

X−Z=1 X−Z=0

没有解决方案。然而，由于假设（3）意味着程序有一个矩阵优化解，它意味着。c属于p×（n+m）的列空间，这里有一个例子，其中admm对一个最优值为−∞的问题发散。

例51.6。考虑以下问题：

f（x）=x，g（z）=0，x−z=0。

由于f（x）+g（z）=x，x=z，变量x不受约束，当x变为−∞时，上述函数变为−∞。增强拉格朗日是

矩阵

-

是奇异的，当（x，z）=t（1,1），t变为−∞时，lρ（x，z，λ）变为−∞in。ADMM步骤如下：

，

这些方程适用于所有k≥0。从最后两个方程我们得出：

，对于所有k≥0，

所以

，对于所有k≥0。

结果我们发现

.

通过归纳，我们得到

，对于所有k≥0，

这表明，当k到无穷大时，xk+3变为−∞，由于xk+2=zk+2，同样地，当k到无穷大时，zk+3变为−∞。

## 51.5停车标准

回到不等式（a2），pk+1−p≤−（λk+1）>Rk+1−ρ（b（zk+1−zk））>（−Rk+1+b（zk+1−z）），（a2）利用ax+bz−c=0和Rk+1=axk+1+bzk+1−c的事实，我们得到

−Rk+1+b（zk+1−z）=−axk+1−bzk+1+c+b（zk+1−z）

=−axk+1+c−bz\_

=−a xk+1+a x=−a（xk+1−x），

所以（a2）可以改写为

pk+1−p≤−（λk+1）>rk+1+ρ（b（zk+1−zk））>a（xk+1−x），

或等同于

pk+1−p≤−（λk+1）>rk+1+（xk+1−x）>ρa>b（zk+1−zk）。（s1）

我们将双重残差定义为

sk+1=ρa>b（zk+1−zk），

数量Rk+1=axk+1+bzk+1−c为主要残余物。那么（s1）可以写成

pk+1−p≤−（λk+1）>rk+1+（xk+1−x）>sk+1。（s）

不等式表明，当残差Rk和Sk较小时，Pk从下到下接近p。因为x是未知的，我们不能使用这个不等式，但是如果我们有一个猜测，那么使用柯西-施瓦兹，我们得到

.

51.6。ADMM的一些应用

上述建议，合理的终止标准是且应该是小的，即

Pri和Dual，

对于某些选定的可行性公差，pri和dual。选择这些参数的进一步讨论可在Boyd等人【28】（第3.3.1节）。

Boyd等人讨论了ADMM的各种扩展和变化。【28】（第3.4节）。为了加速该方法的收敛，可以在每个步骤中选择不同的ρ（例如，ρk），尽管证明这种方法的收敛可能很困难。如果我们假设ρk在多次迭代后变为常数，那么我们给出的证明仍然适用。一个简单的方案是：



K+1=\_\_ρτkincr/τρdecrk ifif

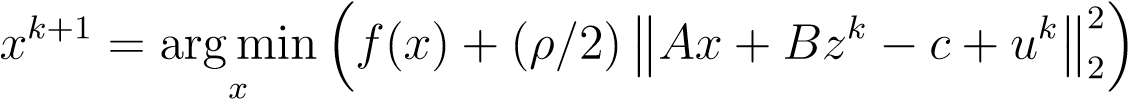
π

\_ρk，否则，

其中，τincr>1、τdecr>1和祄>1是一些选定的参数。我们再次将感兴趣的读者介绍给Boyd等人。【28】（第3.4节）。

## 51.6 ADMM的一些应用

F、G、A和B中的结构经常被用来产生更有效的方法来执行x-update和z-update。我们关注的是x-update，但讨论同样适用于z-update。由于z和λ在x的最小化过程中保持不变，因此使用比例形式的admm更为方便。回想一下



（这里我们使用u而不是μ），因此我们可以将x-update步骤表示为

，

带V=−BZ+C−U。

例51.7。当a=i时出现第一个简化，在这种情况下，x-update是

x+=argminproxf，ρ（v）。

X

map v 7→proxf，ρ（v）被称为f的邻近算子，惩罚为ρ。上述最小化通常称为近端最小化。

例51.8。当函数f足够简单时，可以用解析方法计算邻近算子。特别是当f=ic，非空闭凸集c的指示函数时，这种情况更为明显。

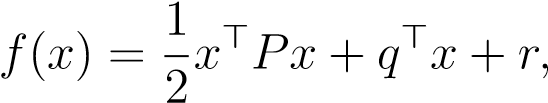
x+=argmin（最小值）

X

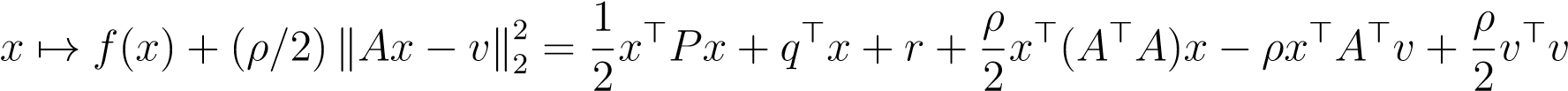
v对c的正交投影。在特殊情况下，其中（第一个或第二个），然后x+=（v）+，

将V的负分量设为零得到的向量。

例51.9。出现简化的第二种情况是f是形式的凸二次函数的情况。



其中p是一个n×n对称的半正定矩阵，q∈rn和r∈r。在这种情况下，图的梯度



由给出

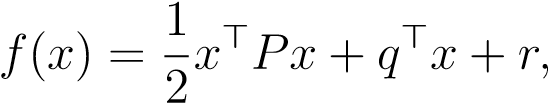
（p+ρa>a）x+q−ρa>v，

由于a的秩为n，所以矩阵a>a是对称正定的，所以我们得到

x+=（p+ρa>a））−1（ρa>v−q）。

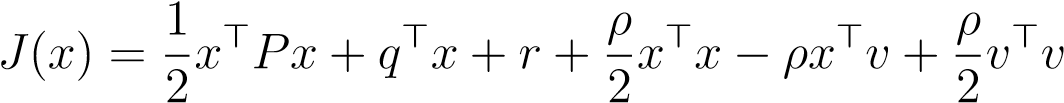
可以使用数值线性代数中的方法来相当有效地计算x+；见Boyd等人。【28】（第4节）。

例51.10。出现简化的第三种情况是前一种情况的变化，其中f是形式的凸二次函数。



除了f受到等式约束cx=b外，如第49.4节所述，这意味着dom（f）=x∈rn cx=b和a=i。x-最小化步骤包括

最小化函数



受约束的

cx=b，

51.6。ADMM的一些应用

因此，根据第49.4节的结果，x+是KKT系统解决方案的组成部分。

.

矩阵p+ρi是对称正定的，因此kkt矩阵是可逆的。

我们现在可以描述如何使用ADMM来解决凸优化的两个常见问题。

（1）在RN中的闭凸集C上，将适当的闭凸集F最小化。这是以下问题

最小化f（x）服从x∈c，

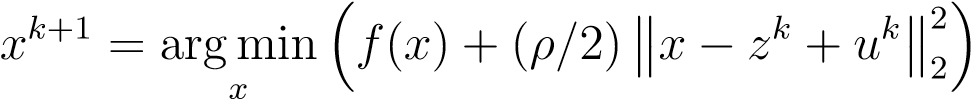
可以用aadm格式重写为

最小化f（x）+ic（z），以x−z=0为准。

使用标度双变量u=λ/ρ，增加的拉格朗日是

.

考虑到示例51.8，此问题的ADMM的缩放形式是



zk+1=\_c（xk+1+uk）

UK+1=UK+XK+1−ZK+1。

X更新包括评估一个近端操作员。注意，函数f不必是可微的。当然，这些最小化依赖于对近端算子和投影算子进行有效的计算。（2）二次规划。问题在这里

减少

以ax=b，x≥0为准，

其中p是一个n×n对称的半正定矩阵，q∈rn，r∈r，a是秩m的m×n矩阵。

上述程序转换为以下ADMM格式：

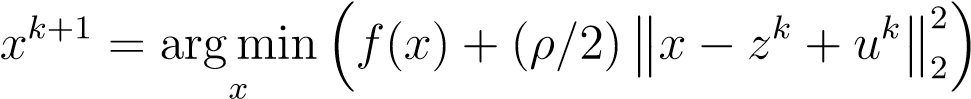
最小化f（x）+g（z），以x−z=0为准，

当dom（f）=x∈rn ax=b，

和

，

正相关函数的指示函数。鉴于实施例51.8和51.10，按比例缩放的ADMM形式包括以下步骤：



zk+1=（xk+1+uk）+

UK+1=UK+XK+1−ZK+1。

X更新涉及求解KKT方程

.

这是一个重要的例子，因为它提供了解决二次问题的最佳方法之一，特别是第54章讨论的SVM问题。

## 51.7 ADMM在1-范数问题中的应用

ADMM的另一个重要应用是“1-范数最小化问题，特别是lasso最小化”，在下面和第52.2节中讨论。这涉及到admm的特殊情况，其中f（x）=τkxk1和a=i。特别是在一维情况下，我们需要解决最小化问题：查找

x=argmin（最小值）

X

x，v∈r，和ρ，τ>0。设c=τ/ρ并写

.

x上的f最小等于x上的f最小

g（x）=2c x+（x−v）2=2c x+x2−2xv+v2，

51.7。ADMM在1-范数问题中的应用

相当于最小化

h（x）=x2+2（c x−xv）

超过x。如果x≥0，则h（x）=x2+2（cx−x v）=x2+2（c−v）x=（x−（v−c））2−（v−c）2。

如果v−c>0，即v>c，因为x≥0，函数x 7→（x−（v−c））2的最小值为

=v c>0，否则如果v−c≤0，则函数x 7→（x−（v−c））2的最小值为

如果x≤0，则h（x）=x2+2（−cx−x v）=x2−2（c+v）x=（x−（v+c））2−（v+c）2。

如果v+c<0，即v<−c，因为x≤0，函数x 7→（x−（v+c））2对于x=v+c具有最小值，否则如果v+c≥0，则函数x 7→（x−（v+c））2对于x=0具有最小值。

总之，infx h（x）是v的函数，由

γ

−V C如果V>C

SC（V）=0，如果V≤C

V+C，如果V<−C。

函数sc被称为软阈值运算符。图中所示的sc图

51.5。

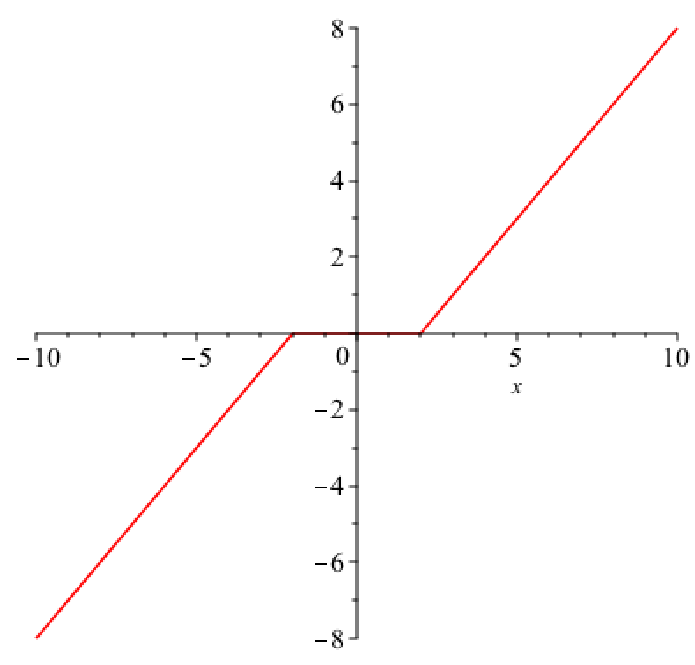


图51.5:sc图（当c=2时）。

你可以查一下

sc（v）=（v−c）+−（−v−c）+，

以及

sc（v）=（1−c/v）+v，v=06，

这表明sc是一个收缩操作符（它向零移动一个点）。

运算符sc扩展到以rn分量表示的向量，即，如果x=（x1，…，xn），则

sc（x）=（sc（x1），…，sc（xn））。

我们现在考虑几个1-范数问题。

1. 最小绝对偏差。

这是最小化kax−bk1而不是kax−bk2的问题。最小绝对偏差比最小二乘拟合更具鲁棒性，因为它能更好地处理异常值。问题可以用ADMM形式表述如下：

最小化kzk1，以ax−z=b为准，

f=0，g=k1。和往常一样，我们假设a是n阶的m×n矩阵，因此a>a是可逆的。ADMM可以表示为

xk+1=（a>a）−1a>（b+zk−uk）zk+1=s1/ρ（axk+1−b+uk）uk+1=uk+axk+1−zk+1−b。

1. 基础追求。

这是以下最小化问题：

最小化kxk1，以ax=b为准，

其中a是秩m<n的m×n矩阵，b∈rm，x∈rn。问题是找到一个欠定线性系统的稀疏解，这意味着有许多零坐标的解X。这个问题在压缩传感和统计信号处理中起着核心作用。

基础追求可以用ADMM形式表示为问题

最小化IC（x）+KZK1，以x−z=0为准，

51.7。ADMM在c=x∈rn ax=b 1-范数问题中的应用。很容易看出，ADMM过程是

xk+1=\_c（zk−uk）

zk+1=s1/ρ（xk+1+uk）uk+1=uk+xk+1−zk+1，

其中\_c是子空间c上的正交投影。实际上，不难证明xk+1=（i−a>（aa>）-1a）（zk−uk）+a>（aa>）-1b。

在某种意义上，“1-最小化问题”被简化为“2-范数问题”的序列。有一些方法可以提高方法的效率；参见Boyd等人。[28]（第6.2节）

1. 一般1-规则化损失最小化。

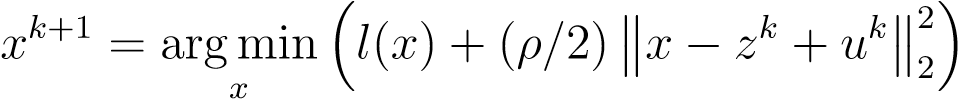
这是以下最小化问题：

最小化l（x）+τkxk1，

式中，L是任何适当的闭和凸损失函数，τ>0。我们将问题转换为ADMM问题：

最小化L（x）+τkzk1，以x−z=0为准。

ADMM程序是



zk+1=sτ/ρ（xk+1+uk）

UK+1=UK+XK+1−ZK+1。

X-update是近端操作员评估。一般来说，我们需要应用一个数值程序来计算xk+1，例如，牛顿方法的一个版本。

特别的情况在哪里特别重要。

（4）套索的正规化。

这是以下最小化问题：

最小化（1.

这是一个线性回归，正则化项τkxk1而不是τkxk2，以鼓励稀疏解。这种方法最初是由Tibshirani arround于1996年提出的，其名称为lasso，代表“最小绝对选择和收缩运算符”。这种方法也被称为“1-正则回归”，但它不如主要使用的“lasso”那么可爱。这种方法在黑斯提、提比西拉尼和温赖特[88]中有广泛讨论。

Lasso最小化转换为以下ADMM形式的问题：

最小化以x−z=0为准。

那么ADMM程序是

xk+1=（a>a+ρi）−1（a>b+ρ（zk−uk））zk+1=sτ/ρ（xk+1+uk）uk+1=uk+xk+1−zk+1。

由于ρ>0，矩阵a>a+ρi是对称正定的。注意，x-update看起来像一个岭回归步骤（参见第52.1节）。

套索有各种各样的概括。

1. 广义拉索正则化。

这是以下最小化问题：

最小化（1，

其中a是m×n矩阵，f是p×n矩阵，a有n秩或f有n秩，该问题转化为admm问题。

减少

以fx−z=0为准，

相应的ADMM程序是

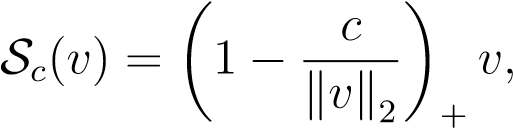
xk+1=（a>a+ρf>f）−1（a>b+ρf>（zk−uk））zk+1=sτ/ρ（fxk+1+uk）uk+1=uk+fxk+1−zk+1。

1. 集体套索。

这是（3）的概括。在这里，我们假设X是X=（x1，…，xn），用Xi＝RNI和N1+Fo.+Xn= n分裂，正则项KXK1被替换。

51.8。总结

. 当ni=1时，这减少到（3）。需要修改admm过程的z-update。我们定义了软阈值算子sc:rm→rm，由



当sc（0）=0时。然后z-update由n个更新组成

zik+1=sτ/ρ（xki+1+uk），i=1，…，n.

该方法可以扩展到处理重叠组；参见Boyd等人。【28】（第6.4节）。

在Boyd等人的研究中，还有许多关于ADMM的应用。[28]，包括共识和分享。另请参见Strang[166]了解简要概述。

## 51.8总结

本章的主要概念和结果如下：

* 双重提升。
* 增强拉格朗日。
* 惩罚参数。
* 乘数法。
* ADMM（乘法器交替方向法）。
* x-更新，z-更新，λ-更新。
* 按比例缩放的ADMM形式。
* 残余，双重残余。
* 停止标准。
* 近端操作员，近端最小化。
* 二次规划。
* KKT方程。
* 软阈值运算符。
* 收缩操作员。
* 最小绝对偏差。
* 基础追求。
* 一般1-规则化损失最小化。
* 套索正规化。
* 广义拉索正则化。
* 集体套索。

三件大波妺

机器学习的应用

一千七百五十一

第五十二章

# 岭回归和套索回归

## 52.1岭回归

解一个超定或欠定线性系统ax=y的问题，是一个“学习问题”，在这个问题中，我们观察一系列数据（（a1，y1），…，（am，ym）），其中ai∈rn和yi∈r，被看作是我们试图推断的未知函数f的输入输出对。最简单的函数是一个线性函数f（x）=x>w，其中w∈rn是一个系数向量，通常称为权向量。由于这个问题是超定的，而且我们的观测可能会有误差，我们不能精确地解W，因为系统的解aw=y，所以我们只能解最小平方问题。

.

在第21.1节（第一卷）中，我们证明了这个问题可以用伪逆法来解决。我们知道极小值w是正态方程a>aw=a>y的解，但是当a>a不可逆时，这种解是不唯一的，因此必须用一些准则在这些解中进行选择。

一种解决方案是选择最小欧几里得范数kw+k2的唯一向量w+，使其最小化。解w+由w+=a+b给出，其中a+是a的伪逆矩阵。矩阵a+由a的SVD得到，即a=v∑u>。也就是说，a+=u∑+v>，其中∑+是用σi−1替换∑中的每一个非零奇异值σi得到的矩阵，保留所有零的位置，然后转置。这种方法的困难在于，它需要知道奇异值是零还是非常小但非零。一个非常小的非零奇异值σin∑会产生一个非常大的值σ−1 in∑+但σ=0在∑中保持为0。

这种不连续现象是不可取的，另一种方法是通过在kaw−yk2中添加一个正则化项来控制w的大小，自然的候选项是kwk2。还可以将矩阵A的每一行视为输入向量Xi

一千七百五十三

将m×n矩阵x定义为

，

其中行向量x> i是x的行，因此Xi→rn是列向量。我们的优化问题称为岭回归，即问题（rr1）：

最小化KY−XWK2+K KWK2，

通过引入新变量ξ=y−xw，可以重写为（rr2）：

最小化ξ>ξ+kw>w取决于

y−xw=ξ，

其中k>0是确定正则化项w>w影响的常数。

我们的最小化问题的第一版的目标函数可以表示为

j（w）=ky−xwk2+k kwk2

=（y−x w）>（y−xw）+kw>w=y>y−2w>x>y+w>x>xw+kw>w=w>（x>x+kin）w−2w>x>y+y>y。

矩阵x>x为对称半正定，k>0，因此矩阵x>x+kin为正定。接下来是

j（w）=w>（x>x+kin）w−2w>x>y+y>y

是严格凸的，所以它有一个独特的最小iff jw=0。自从

jw=2（x>x+kin）w−2x>y，

我们推断

W=（x>x+kin）−1X>y.（wp）

有趣的是，当k>0变为零时，矩阵（x>x+kin）−1X>的极限是x的伪逆x+。为了证明这一点，让x=v∑u>是x的SVD。

然后

（x>x+kin）=u∑>v>v∑u>+kin=u（∑>∑+kin）u>，

所以

（x>x+kin）−1X>=u（∑>kin）−1U>u∑>v>=u（∑>kin）−1∑>v>。矩阵中的对角项（∑>∑+kin）−1∑>是

，如果σi>0，

如果σi=0，则为零。所有非术语项均为零。当σi>0和k>0变为0时，

，

所以

lim（∑>∑+kin）−1∑>=∑+，

K7～0

这意味着

lim（x>x+kin）−1x>=x+。

K7～0

我们问题的第一个公式的对偶函数是一个常数函数（其值最小为j），因此它是无效的，但是我们问题的第二个公式产生了一个有趣的对偶问题。拉格朗日是

L（ξ，w，λ）=ξ>ξ+kw>w+（y−xw−ξ）>λ

=ξ>ξ+kw>w−w>x>λ−ξ>λ+λ>y。

用λ，ξ，y∈rm。

为了推导对偶函数g（λ），我们将l（ξ，w，λ）对ξ和w最小化，为此我们将梯度lξ，w设为零。自从

，

我们得到

.

上面建议定义变量α，使ξ=kα，因此我们得到λ=2kα和w=x>α。然后用拉格朗日方程中的ξ、λ和w的上述值代入α的函数，得到了α的对偶函数。

g（α）=k2α>α+kα>xx>α−2kα>xx>α−2k2α>α+2kα>y=−kα>（xx>+kim）α+2kα>y。

这是一个严格的凹函数，因此它的最大值达到iff gα=0，即，

2k（xx>+kim）α=2ky，

会产生

α=（xx>+kim）−1y.

把我们所获得的一切结合起来

α=（xx>+kim）−1y

w=x>α

ξ=kα，

会产生

w=x>（xx>+kim）−1y.（wd）

在早期（wp）我们发现

W=（X>X+K）−1X>Y，

而且很容易检查

（x>x+kin）−1X>=x>（xx>+kim）−1.

采用上述方法学习仿射函数f（w）=x>w+b，而不是线性函数f（w）=x>w，其中b∈r，我们有以下优化程序（rr3）：

最小化ξ>ξ+kw>w

从属于

y−xw−b1=ξ，

其中y，ξ，1∈Rm，w∈Rn。注意，在程序（rr3）中，最小化仅在ξ和w上执行，而不是在变量b上执行。与此程序相关的拉格朗日是

L（ξ，w，b，λ）=ξ>ξ+kw>w−w>x>λ−ξ>λ−b1>λ+λ>y。

通过将梯度lξ，b，w设置为零，我们得到

.

如前所述，如果我们设置ξ=kα，我们得到w=x>α和

g（α）=-kα>（xx>+kim）α+2kα>y。

由于k>0且λ=2kα，双岭回归为以下程序（drr3）：

最小化α>（xx>+kim）α−2α>y

1>α=0.

注意，在系数1/2之前，这个问题满足命题41.3的条件，a=（x x>+kim）−1，b=y，b=1m，f=0，x重命名为α。因此，它有一个独特的解α（注意，λ=2kα不是命题41.3中使用的λ，我们将其重命名为μ）。因为命题41.3给出的解是

µ=（b>ab）−1（b>ab−f），α=a（b−b）

我们得到μ=（1>（xx>+kim）−11>（xx>+kim）−1y，α=（xx>+kim）−1（y−1）。

注意，矩阵b>ab是标量1>（xx>+kim）−11。

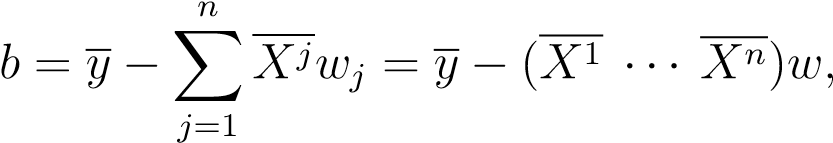
一旦α，ξ=kα，w=x>α被确定，b由方程给出

b1=y−xw−ξ=y−xw−kα。

由于1>1=m和1>α=0，我们得到

，

其中y是y的平均值，xj是x的jth列的平均值。因此，



其中（x1···xn）是1×n行向量，其jth项为xj。因为w=x>α，我们可以

同时写

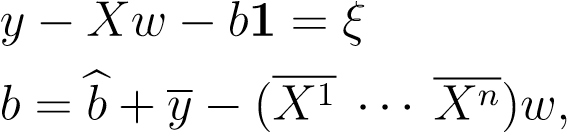
表达式

-。···

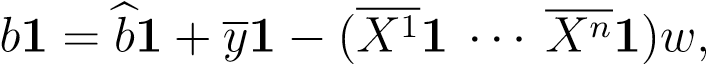
建议寻找上述形式的截距项b（也称为偏差），即程序（rr4）：

最小化ξ>ξ+kw>w

从属于



同样，在程序（rr4）中，最小化只在ξ和w上进行，因为



如果x=（x11···xn1）是m×n矩阵，其jth列为矢量xj1，则上述程序等于程序（rr5）：

最小化ξ>ξ+kw>w取决于

如果我们写y=y−y1和x=x−x，则上述程序变为（rr6）：

乙

根据Yb−xwb−bb1=ξ，最小化ξ>ξ+kw>w。

如果该程序的解决方案是wb，则bb由

bb=yb−（xb1···xbn）wb=0，

因为数据Yb和Xb居中。因此（rr6）相当于岭回归

没有适用于中心数据的截距项yb=y−y1和xb=x−x，程序（rr60）：

根据Yb−xwb=ξ，最小化ξ>ξ+kw>w。

如果w是由b给出的该程序的最优解

wb=xb>（xbxb>+kim）−1y，b

那么b是由

B=Y−（x1···xn）W.

乙

备注：虽然这不是一个明显的先验，但程序（rr3）的最优解w等于程序（rr60）的最优解w。然而，实际上，自从

B求解对偶（drr3）比求解程序（rr60）困难，因为对偶程序具有额外的约束1>α=0，所以涉及中心数据的程序（rr60）是首选的。

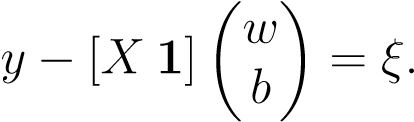
如果我们在程序（rr3）中也将b最小化，自然会想知道会发生什么。让我们将术语KB2添加到目标函数中。然后我们得到程序

根据y−xw−b1=ξ，最小化ξ>ξ+kw>w+kb2。

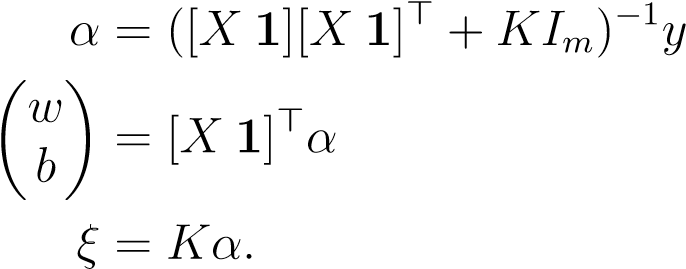
这建议将b作为权重向量w的一个额外分量，通过将m×n+1矩阵[x 1]的列（维度m的列）添加到矩阵x中，得到程序（rr3b）：

最小化ξ>ξ+kw>w+kb2

从属于



这个程序和程序（rr2）一样被解决，我们得到



因此b=1>α。

观察[x 1][x 1]>=xx>+11>。因为我们也有这个方程

y−xw−b1=ξ，

我们得到

所以

会产生

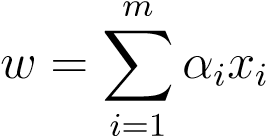
.

用任意常数c>0重新计算k的推导完全相同，我们得到

.

正如黑斯提、蒂比西拉尼和弗里德曼[87]所指出的那样（第3.4节），方法的一个缺陷是B的解在给每个值yi加上一个常数c时不是不变的。使用程序（RR60）的方法并非如此。

对偶（rr2）或（rr3）的一个有趣的方面是，它表明形式x>α的解w是一个线性组合。



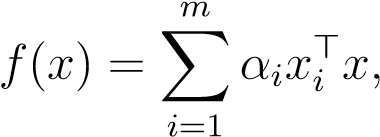
在数据点Xi中，系数αi对应于对偶函数的双变量La2= 2kα，以及

α=（xx>+kim）−1y.

如果m小于n，则求解α更为有利。但真正让双重意义有趣的是，我们对x的定义是

，

矩阵x x>由内部积x>i xj组成，同样，函数学习f（x）=w>x可以表示为



即w和f（x）都是根据内积x>i xj和x>i x给出的。

这一事实是岭回归推广的关键，在岭回归中，输入空间rn嵌入到更大（可能是无限维）的欧几里得空间f（内积h−、−i）中，通常使用函数称为特征空间。

⑨：Rn→F.

问题变成（核岭回归）（krr2）：

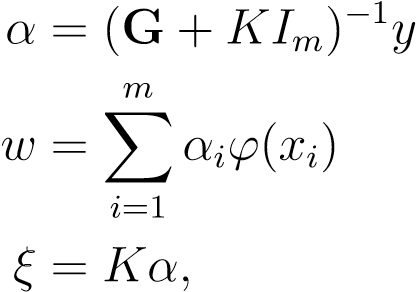
最小化ξ>ξ+khw，wi

以Y-HW为例，（i）i＝ZeI i，i＝1，…，m。

注意w∈f。这个问题在Shawe–Taylor和Christianini[154]中讨论（第

7.3）。

我们将在下面说明解决方案完全相同：



核矩阵，其中G是给定的GM矩阵，并且通常用GOJJ＝H\*（Xi）G，，（XJ）I表示。

K

在这个框架中，我们在使用梯度时必须小心谨慎，因为涉及到内积h−、−i on f，而f可以是无限维的，但这不会造成任何问题，因为我们可以使用导数，并且根据命题38.5，我们已经

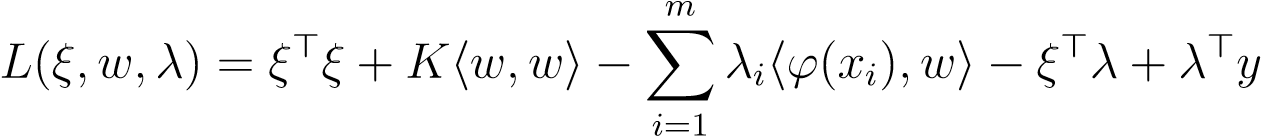
dh−、−i（u，v）（x，y）=hx，vi+hu，yi。

这意味着地图u 7→hu，ui的导数是

dh−、−iu（x）=2hx，ui。

由于地图u 7→hu，vi（v固定）是线性的，其导数为

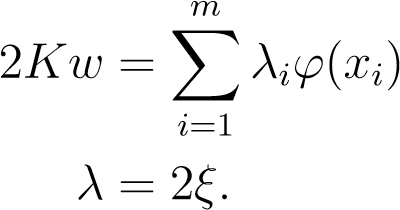
dh−，viu（x）=hx，vi.拉格朗日的导数



关于ξ和w是

.

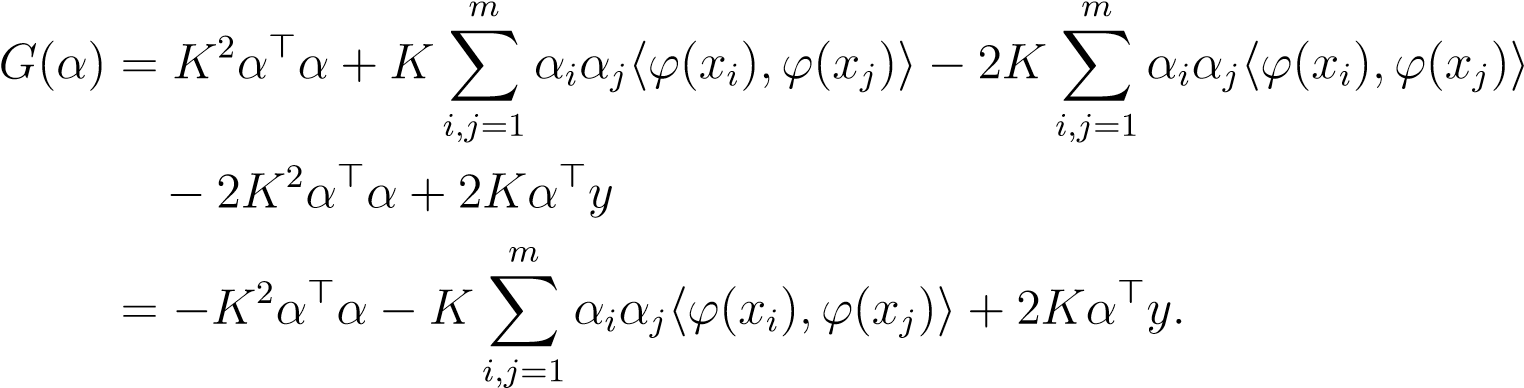
我们有=0表示全部，如果



我们再次定义了ξ=kα，所以我们得到了λ=2kα，并且

.

回到拉格朗日



如果G是由Gij= H\*（Xi），（xJ）i给出的矩阵，那么我们得到

G（α）=-Kα>（G+Kim）α+2Kα>Y。

函数g是严格凹形的，其最大值为

α=（g+kim）−1y，

如前所述。

在岭回归的标准情况下，如果f=rn（但内积h−、−i是任意的），我们可以利用上述方法学习仿射函数f（w）=x>w+b，而不是线性函数f（w）=x>w，其中b∈r。这次我们假设b是形式

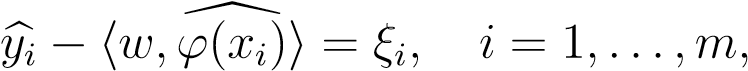
b=y−hw，（x1···xn）i，

其中x j是m×n矩阵x的j列，其第i行是

列向量ε（Xi），其中（x1·…·xn）被视为列向量。我们有最小化问题（krr60）：

最小化ξ>ξ+khw，wi

从属于



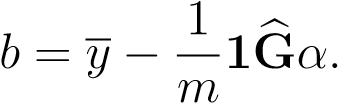
其中）是n维向量

根据矩阵G给出的解由

，

和以前一样。我们得到α=（gb+kim）−1y，b

根据前面的计算，b由



我们将在第53.3节中解释如何从矩阵G计算矩阵gb。

由于特征空间F的维数可能非常大，人们可能担心计算内积H（Xi），（xJ）i可能非常昂贵。这就是内核函数的救星。嵌入的一个核函数κ：rn→f是一个mapκ：rn×rn→r，其性质是

κ（u，v）=h（u），（v）i表示所有u，v∈rn。

如果可以以相当便宜的方式计算κ（u，v），并且如果（u）可以便宜地计算，那么可以很便宜地计算内积H\*（Xi）、（xJ）i（和h（x））、（x）i。幸运的是，有好的内核函数。关于内核方法的两个非常好的来源是scho–lkopf和smola[141]以及shawe–taylor和Christianini[154]。我们将在第53章中研究内核。

## 52.2套索回归（`1-正则化回归）

岭回归的主要缺点是估计权向量w通常具有许多非零系数。因此，岭回归不能很好地扩大。实际上，我们需要能够处理数百万个或更多参数的方法。一种鼓励向量w稀疏的方法是用惩罚函数k kwk1替换二次惩罚函数，用“1-范数”替换“2-范数”。

这种方法最初是由Tibshirani arround于1996年提出的，其名称为lasso，代表“最小绝对选择和收缩运算符”。这种方法也被称为“1-正则回归”，但它不如主要使用的“lasso”那么可爱。

给定一组训练数据{（x1，y1），…，（xm，ym）}，用xi rn和yir r，如果x是m×n矩阵

，

其中，行向量x>i是x的行，如果出现以下优化问题（lasso1），则lasso回归：

最小化

y−xw=ξ，

其中k>0是确定正则化项kwk1影响的常数。

规则化术语kwk1=w1+··················wn的困难在于，对于所有w，映射w 7→kwk1都是不可微的。这种困难可以通过使用子梯度来克服，但也可以通过使用我们已经掌握的技巧，以基本的方式获得上述程序的对偶。使用，即如果x∈r，则

| X=最大X，−X。

利用这个技巧，通过引入一个非负变量的向量，我们可以重写lasso最小化，如下所示：lasso正则化（lasso2）：

最小化

，

其中y，ξ∈rm和。

约束和是等价的，对于一个最优解，我们必须有，也就是。

拉格朗日）由

，

与λ∈Rm和。由于目标函数是凸的，且约束L相对于原始函数具有最小值为仿射（因此是合格的），所以拉格朗日

变量，=0。因为梯度由

，

我们得到了方程

ξ=λ

α+α-α-β=x>λα+α-β=k1-β。

利用这些方程，给出了

.

由于β≥0，约束α+α−=k1−β等于

α+α−≤k1。

其中最大值为α+，α−≥。如果我们回想一下，对于任意i∈1，…，n zt（∈rn，α+）i−（α-）i的最小值是−k，并且

K

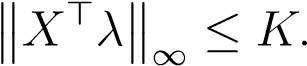
kzk∞=1最大值≤i≤n zi，

因此，约束条件

α+α−≤k1

x>λ=α+α−α−

相当于

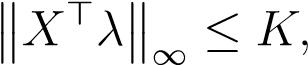


以上相当于2n约束

−k≤（x>λ）i≤k，1≤i≤n。

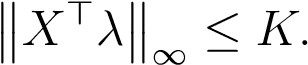
因此，双套索程序由

最大化受试者



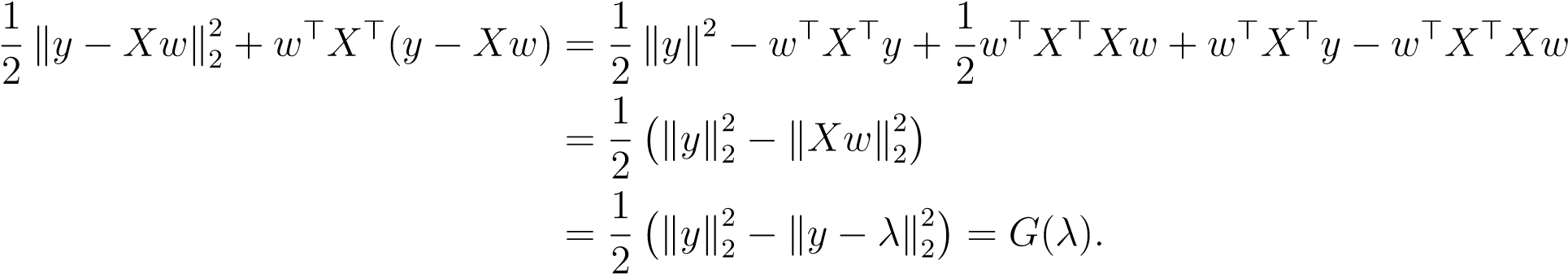
其中（因为是常数项）等于（dlasso2）：

最小化



考虑到约束y−xw=ξ，并且对于最优解，我们必须有ξ=λ，必须满足以下条件：

（）还观察到，对于最优解，我们



由于目标函数是凸的，且常数是合格的，对偶间隙为零，因此对于初等和对偶的最优解，即

，

由此得出方程

w>x>（y−xw）=k kwk1.（）

以上是w和x>的内积（y−xw），所以当w i=06时，由于kwk1=w1+············································如果

S=I∈1，…，N Wi=06，

如果x表示由s索引的x列组成的矩阵，如果ws表示由w的非零分量组成的向量，那么我们有

.

我们也有

−∞

其中s是s的补码。第一个方程得出

xs>xs ws=xs>y-ksgn（ws）

因此，如果x>xs是可逆的（如果x的列是线性无关的，则是这种情况），我们得到ws=（xs>xs）−1（xs>y-ksgn（ws））。

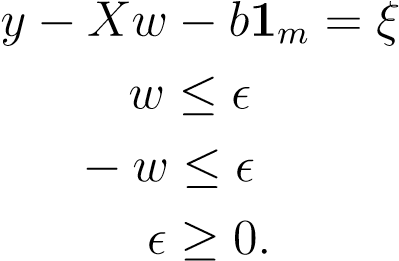
理论上，如果我们知道w的支持和它的组件的符号，那么ws是确定的，但在实践中，这是无用的，因为问题是要找到支持和解决方案的符号。

求解套索回归的一种方法是用对偶程序求出λ=ξ，然后用线性规划方法求出W，通过保持ξ常数求解套索初值产生的线性规划。最好的方法是使用第51.7（5）节所述的ADMM。梯度下降也有许多变化；见黑斯提、提比西拉尼和温赖特。

〔88〕。

在前面的讨论中，我们假设我们试图学习一个线性函数f（x）=w>x。为了学习仿射函数f（x）=w>x+b，我们解决了以下优化问题（lasso3）：

最小化



观察到，正如在岭回归的情况下，我们没有最小化超过b。

与这个优化问题有关的拉格朗日是

，

因此，通过将梯度设置为零，我们得到了方程

ξ=λ

α+α-α-β=x>λα+α-β=k1-β

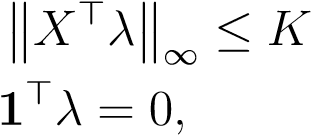
1>λ=0，

利用这些方程，我们发现对偶函数也由

，

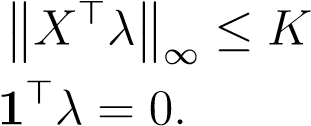
双套索程序由

最大化受试者



相当于（dlasso3）：

最小化



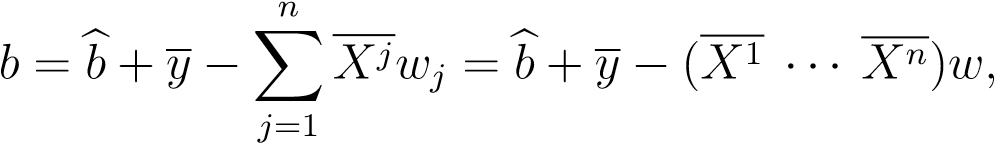
一旦确定了λ=ξ和w，我们就用方程得到b。

b1=y−xw−ξ，

由于1>1=m和1>ξ=1>λ=0，上述结果

，

其中y是y的平均值，xj是x的jth列的平均值。



可用于，如岭回归（见第52.1节），证明程序（lasso3）等同于对中心应用无截距项的lasso回归（lasso2）。

数据，替换b。那么b是由

其中w是（lasso2）给出的溶液。这是黑斯迪，蒂比西拉尼描述的方法，

乙

和Wainwright[88]（第2.2节）。

找到b的另一种方法是将术语（c/2）b2添加到目标函数中，使某个正常数c获得程序（lasso4）。这次拉格朗日是

，

因此，通过将梯度设置为零，我们得到了方程

ξ=λ

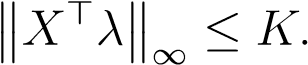
α+α-α-β=x>λα+α-β=k1-β

cb=1>λ。

52.3。总结

因此B也被确定，并且双套索程序与第一套索双（dlasso2）相同，即

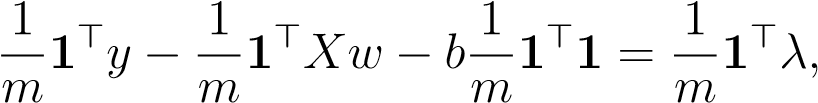
最小化



因为方程ξ=λ和

y−xw−b1=ξ

保持，从cb=1>λ我们得到



那就是

会产生

.

在岭回归的情况下，B也受到惩罚的方法的一个缺陷是，在每个值yi加上一个常数c时，B的解不是不变的。

## 52.3总结

本章的主要概念和结果如下：

* 岭回归。
* 核岭回归。
* 内核函数。
* 套索回归。

1770年第52章。岭回归和套索回归

第五十三章

# 正定核

## 53.1正定核的基本性质

设x为非空集。如果集合x代表一组高度非线性的数据，使用一个称为特征图的函数，将x映射到一个更高维度的空间h，这可能是有利的。这一观点是，将x中的对象描述“展开”，以使其线性化。空间h通常是一个向量空间，它有一个内积h−−i。如果h是无限维，那么我们假设它是希尔伯特空间。

许多分析或分类数据的算法都使用了内积h\_（x），（y）i，其中x，y∈x。因此，自然会做出以下定义。

定义53.1.设x为非空集，设h为（复杂）希尔伯特空间，设\_：x→h为称为特征图的函数。函数k:x×x→c由

κ（x，y）=h（x），（y）i，x，y∈x，

称为内核函数。

备注：一个特征映射通常被称为特征嵌入，但是这个术语有点误导，因为它表明这样的映射是可注入的，这不一定是如此。不幸的是，大多数人都使用这个术语。

例53.1。假设我们有两个特征图，分别是：\_:x→rn1和\_:x→rn2，让κ1（x，y）=h\_（x），\_（y）i和κ2（x，y）=h\_（x），\_（y）i是对应的核函数1（其中h−，−i是rn上的标准内积）。定义特征图\_：x→rn+n2 by\_（x）=（\_1（x），\_2（x）），

（n1+n2）元组。我们有h\_（x），\_（y）i=h（\_1（x），\_（x）），（\_（y），\_（y））i=h\_（x），\_（y）i+h\_（x），\_（y）i

=κ1（x，y）+κ2（x，y），

一千七百七十一

它显示了由

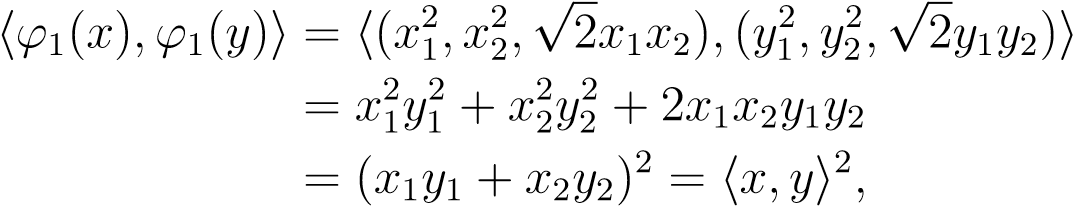
K（x，y）=K 1（x，y）+K 2（x，y）

是对应于特征映射的核心函数，即：x→rn1+n2。

例53.2。设x为r2的一个子集，设\_:x→r3为下式给出的图

.

观察特征空间h=r3中的线性关系对应于（数据）输入空间中的二次关系。我们有



其中hx，yi是r2上常见的内积。因此函数

κ（x，y）=hx，yi2

是与功能空间r3关联的内核函数。如果我们现在考虑图\_:x→r4，由

，

我们马上检查

h\_2（x），\_2（y）i=κ（x，z）=hx，yi2，

这表明相同的内核可以从不同的映射中产生到不同的特征空间。

例53.3。实例53.2可概括如下。假设我们有一个特征图\_:x→rn，让κ1（x，y）=h\_（x），\_（y）i是相应的核函数（其中h−，−i是rn上的标准内积）。通过其N2组分（x）（i，j）=（\_（x））i（\_（x））j，1≤i，j≤n，定义特征图\_：x→rn×rn。

内积在Rn×Rn上，由

.

然后我们有了



因此，由κ（x，y）=（κ1（x，y））2给出的mapκ是一个与特征图\_：x→rn×rn相关联的核映射。特征图a是实例53.2中特征图\_的直接概括。

2.上面的论点立即被改编以表明，如果\_:X→RN1和\_:X→RN是两个特征映射，并且如果κ1（x，y）=h\_（x），\_（y）i和κ2（x，y）=h\_（x），\_（y）i是对应的核函数，那么由定义的映射

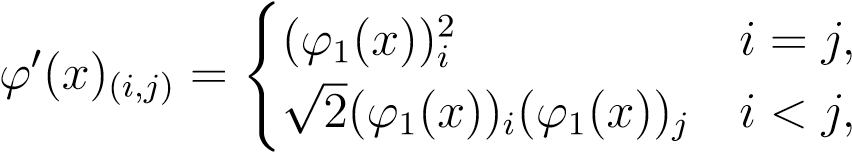
k（x，y）=k 1（x，y）k 2（x，y）

是一个核心函数，用于特征空间RN1×RN2和特征映射

⑨（x）（i，j）=（\_（x））i（\_（x））j，1≤i≤n1，1≤j≤n2.

例53.4。注意，特征图\_：x→rn×rn不是很经济的，因为如果i=6 j，那么部件\_（i，j）（x）和\_（j，i）（x）都等于（\_1（x））i（\_1（x））j。

因此，我们可以定义由



其中1≤i≤j≤n的对（i，j）按字典顺序排列。特征图a是实例53.2中特征图\_的直接概括。

请注意，也可以用以下方式定义\_，这样更容易得出任何幂的泛化：

，

其中n-元组（i1，…，in）按字典顺序排列。回想一下，对于任意m≥1和任意（i1，…，in）∈nm，这样i1+i2+······+in=m，我们有

.

更一般地，对于任何m≥2，利用多项式定理，我们可以定义一个特征嵌入，定义由κ（x，y）=（κ1（x，y））m给出的核函数κ，其中

，

其中n-元组（i1，…，in）按字典顺序排列。

例53.5。对于任何正实常数r>0，常数函数k（x，y）=r为

一个对应于特征图的核函数，由\_（x，y）=√r给出。

根据定义，给出的函数是一个核函数。

（特征图是从RN到自身的标识图）。我们刚刚看到，对于任何正实常数r>0，该常数都是一个核函数。例53.1中，函数）是一个核函数，对于任何整数d≥1，by

例53.3，函数k d由

，

是RN上的内核函数。根据二项式，

.

通过示例53.1，该内核函数的特征映射是d+1内核映射rd−mhx、yim的特征的串联。通过示例53.3，内核映射rd−mhx，yim的特征映射的组件是函数的重加权。



用（I1，…，in）∈nn。因此，核函数κd的特征图的组成部分是函数的重加权。



用（I1，…，in）∈nn。很容易看出这个特征空间的尺寸是。

多项式核κD有许多变化；所有嵌入核的子集、方差分析核；见Shawe–Taylor和Christianini[154]，第三章。

在下一个示例中，集合X不是向量空间。

例53.6。设d为有限集，x=2d为其幂集。如果d\_=n，设h=rx～r2n。我们假设以某种方式枚举d的子集，使r2n的每个坐标对应于其中一个子集。例如，如果d=1,2,3,4，则让

U1=∅U2=1 U3=2 U4=3

U5=4 U6=1,2 U7=1,3 U8=1,4

U9=2,3 U10=2,4 U11=3,4 U12=1,2,3

U13=1,2,4 U14=1,3,4 U15=2,3,4 U16=1,2,3,4。

设\_：x→h为特征图，定义如下：对于任意子集a，u∈x，

（

1如果U A

⑨（a）U=

否则为0。

例如，如果a1=1,2,3，我们得到向量

（1,2,3）=（1,1,1,1,0,1,0,1,0,0,1,0,0,0,0,0,0,0,0），

如果a2=2,3,4，我们得到矢量

（2,3,4）=（1,0,1,1,1,1,0,0,0,1,1,0,0,0,1,0）。

对于D的任意两个子集A1和A2，很容易检查

h\_（a1），（a2）i=2 a1 a2，

a1和a2的公共子集的数目。例如，A1 A2=2,3，和

h\_（a1），\_（a2）i=4.

因此，函数k:x×x→r由

κ（a1，a2）=2 a1 a2，a1，a2 d

是一个内核函数。

内核函数具有以下重要特性。

提案53.1.设x为任意非空集，设h为任意（复杂）希尔伯特空间，设\_：x→h为任意函数，设k:x×x→c为核，由

κ（x，y）=h（x），（y）i，x，y∈x.

对于x的任何有限子集s=x1，…，x p，如果ks是p×p矩阵

KS＝（κ（Xj，Xi））1，i，j，ω＝p＝（H\*（Xj），（Xi）i）1，i，j，ωp，

然后我们得到u ks u≥0，对于所有u cp。

证据。我们有

，

如要求。

命题53.1提出了对内核函数的第二种方法，该方法不假定提供了特征空间和特征映射。我们将在第53.2节中看到这两种方法是等效的。第二种方法在实践中很有用，因为通常很难以简单的方式定义特征空间和特征映射。

定义53.2.设x为非空集。对于x的每个有限子集s=x1，…，x p，如果ks是p×p矩阵，那么函数k:x×x→c是一个正定核。

KS＝（κ（Xj，Xi））1，i，j，ωp

我们称之为克矩阵

，对于所有u∈cp。

注意，如果u=06，定义53.2不要求u ks u>0，因此术语正定有点滥用，使用术语正半定更合适。然而，似乎习惯上使用“正定核”，甚至是“正定核”。

提案53.2.设k:x×x→c为正定核。然后，对于所有x∈x，k（x，x）≥0，对于x的任何有限子集s=x1，…，x p，p×p矩阵k由

KS＝（κ（Xj，Xi））1，i，j，ωp

是赫米提安，也就是说，ks=ks。

证据。第一个属性通过选择s=x很明显。我们有

（u+v）ks（u+v）=u ksu+u ksv+v ksu+v ksv，

|  |  |
| --- | --- |
| 由于（u+v）ks（u+v），u ksu，v ksv≥0，我们推断 |  |
| 2a=u ksv+v ksu  一定是真的。用IU代替U，我们看到了 | （1） |
| 2b=−iu ksv+iv ksu | （2） |

也必须是实数，通过将方程（2）乘以i并将其加到方程（1）中，我们得到

u ksv=a+ib.（3）

从式（1）中减去式（3），我们得到

然后

-

对于所有u，v∈c，这意味着ks=ks。

如果mapκ：x×x→r是实数，那么我们有如下的标准，即κ是一个只涉及实数向量的正定核。

提案53.3.如果k:x×x→r，那么对于任何有限子集s x1，…，x p of x，p×p实矩阵k，由下式给出

ks=（k（xk，xj））1≤j，k≤p

是对称的，即，和

，对于所有u∈rp。

证据。如果κ是实值正定核，那么这个命题就是53.2命题的一个微不足道的结果。

相反，假设κ是对称的，并且它满足命题的第二个条件。我们需要证明，对于复向量，κ是一个正定核。如果我们写UK=AK+IBK，那么

.

因此，u ksu是真实的，iff ks是对称的。

因此，我们做出以下定义。

定义53.3.如果k（letx，y）=x是一个非空集，则确定核。一个函数k（y，x）对于所有x，y∈x，对于每个有限子集，k:x×x→r是x的（real）正x1，…，x p，如果ks是p×p实对称矩阵

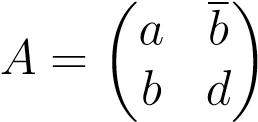
KS＝（κ（Xi，Xj））1，i，j，ωp，

然后我们有了

，对于所有u∈rp。

除其他外，下一个命题表明一个正定核满足柯西-施瓦兹不等式。

提案53.4.厄米提亚2×2矩阵



当且仅当a≥0、d≥0和ad−b 2≥0时为正半定。

设k:x×x→c为正定核。对于所有x，y∈x，我们有κ（x，y）2≤κ（x，x）κ（y，y）。

证据。对于所有的x，y∈c，我们有

.

如果a是半正定的，那么我们已经知道a≥0和d≥0。如果a=0，则

我们必须使b=0，否则我们可以使bxy+bxy，它是bxy的两倍实部，如我们想要的那样为负。在这种情况下，ad−b 2=0。

如果a>0，则

.

因此，如果ad−b 2<ad0−，我们可以选择b 2≥0。相反，y 6=0，x=−（by）/a，所以上面的表达式是负数。

如果x=y，则κ（x，y）2≤κ（x，x）κ（y，y）的不等式是微不足道的。如果x 6=y，则不等式通过将正半定准则应用于矩阵而得到。

，

如要求。

由舒尔（1911）提出的以下性质表明，两个正定核的点积也是正定核。

提案53.5。（I.Schur）如果k 1:x×x→c和k 2:x×x→c是两个正定核，那么由k（x，y）=k 1（x，y）k 2（x，y）给出的函数k:x×x→c对于所有x，y∈x也是一个正定核。

证据。证明了如果a=（ajk）和b=（bjk）是两个厄米半正定p×p矩阵，那么它们的点积c=a b=（ajkbjk）（也称为hadamard或schur积）。回想一下，厄米特半正定矩阵a可以对角化为a=u∧u，其中∧是一个带非负项的对角矩阵，u是一个单位矩阵。让∧1/2是由∧中对角线项的正平方根组成的对角线矩阵。然后我们有了

A=U∧U=U∧1/2∧1/2U=U∧1/2（U∧1/2）。

因此，如果我们设置r=u∧1/2，我们有

A=RR，

也就是说

.

那么对于任何u∈cp，我们有

.

既然b是半正定的，对于每个固定的h，我们有

，

如我们所见，让z=（u1r1h，…，uprph）

相反，两个对称半正定矩阵A和B的常积AB可能不是对称半正定的；示例见第7.9节。

以下是从旧核中获得新的正定核的其他方法。

提案53.6.设k 1:x×x→c和k 2:x×x→c为两个正定核，f:x→c为函数，ψ：x→rn为函数，k 3:rn×rn→c为正定核，a∈r为任意正实。那么下面的函数就是正定核：

.

（5）如果b是一个对称的半正定n×n矩阵，那么mapκ：rn×rn→r由

κ（x，y）=x>

是一个正定核。

证据。（1）对于x的每个有限子集s=x1，…，x p，如果k1是p×p矩阵

k1=（k 1（xk，xj））1≤j，k≤p

如果k2是p×p矩阵

k2=（κ2（xk，xj））1≤j，k≤p，

那么对于任何u∈cp，我们有

u（k1+k2）u=u k1u+u k2u≥0，

因为u k1u≥0和u k2u≥0，因为κ2和κ2是阳性的定核，这意味着k1和k2是阳性的半定核。

1. 我们有u（ak1）u=au k1u≥0，

因为a>0且u k1u≥0。

1. 对于x的每个有限子集s=x1，…，x p，如果k是p×p矩阵

K=（K（XK，XJ））1≤J，K≤P=（F（XK）F（XJ））1≤J，K≤P

然后我们有了

.

1. 对于x的每个有限子集s=x1，…，x p，p×p矩阵k由

k=（k（xk，xj））1≤j，k≤p=（k 3（ψ（xk），ψ（xj）））1≤j，k≤p

是对称的半正定的，因为κ3是一个正定核。

1. 在53.5号提案的证明中（与实际情况相适应），有一个矩阵r

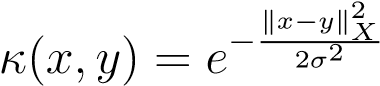
B=RR>，

所以κ（x，y）=x>by=x>r r>y=（r>x）>r>y=hr>x，r>yi，

因此，κ是由特征映射（x）=r>x从rn到自身给出的核函数，根据命题53.1，它是一个对称的正定核。

提案53.7.设k 1:x×x→c为正定核，p（z）为非负系数多项式。然后下面定义的函数k也是正定核。

1. κ（x，y）=p（κ1（x，y））。
2. κ（x，y）=eκ1（x，y）。
3. 如果x是带内积h−、−ix和相应的范数k kx的实希尔伯特空间，



对于任何σ>0。

证据。（1）如果p（z）=amzm+····+a1z+a0，则

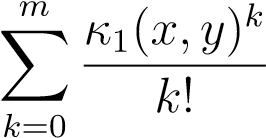
p（k 1（x，y））=am k 1（x，y）m+······+a1 k 1（x，y）+a0。

由于k=0，…，m的ak≥0，根据命题53.5和命题53.6（2），1≤k≤m的每个函数akκ√i（x，y）k是一个正定核，根据命题53.6（3），f（x）=a0，常数函数a0是一个正定核，根据命题53.6（1），p（k 1（x，y））是一个正定核。有组织的定核。

1. 我们有

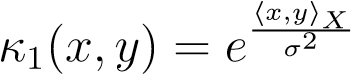
.

（1）部分和



是正定核，由于eκ1（x，y）是正定核的（均匀）点限，所以也是正定核。

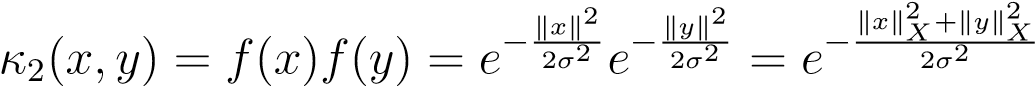
1. 根据命题53.6（2），由于map（x，y）7→hx，yix显然是一个正定核（特征映射是一个恒等式），并且由于σ=06，函数（x，y）7→hx，yix/σ2是一个正定核，因此由（2）可知，



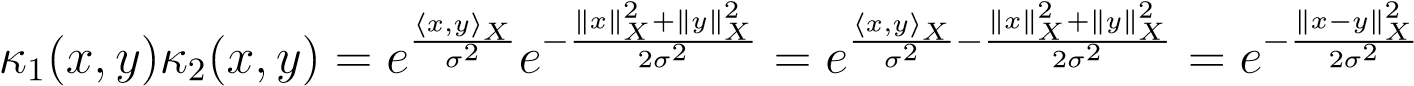
是一个正定核。设f:x→r为

.

然后根据第53.6（3）号提案，

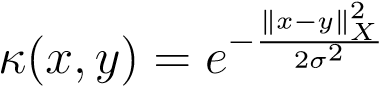


是一个正定核。根据命题53.5，功能κ1κ2是一个正定核，即



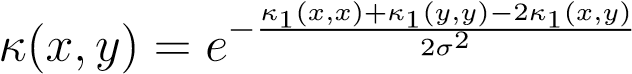
是一个正定核。

正定核



称为高斯核。这个内核需要一个无限维空间中的特征映射，因为它是不同内核的无穷和。

注：如果κ1是一个正定核，则命题53.7（3）的证明立即适用于证明



是一个正定核。

其次，我们证明了每一个正定核都来自希尔伯特空间中的一个特征映射，它是一个函数空间。

## 53.2正定核的希尔伯特空间表示

下面的结果展示了如何从一个正定核构造一个所谓的复制核希尔伯特空间，简称Rkhs。

53.2。正核的希尔伯特空间表示

定理53.8。设k:x×x→c为非空集x上的一个正定核，对于每个x∈x，设k x:x→c为

κx（y）=κ（x，y），y∈x。

设h0为函数族（κx）∈x所跨越的x到c函数的向量空间cx的子空间，设\_：x→h0为（x）=κx给出的映射，h0上有一个厄米特内积H−、−i，这样

κ（x，y）=h（x），（y）i，对于所有x，y∈x。

h0的完成h是希尔伯特空间，图η：h→cx由

ε（f）（x）＝Hf，κXi，x x，

是线性的和内射的，所以h可以用cx的子空间来标识。我们也有

κ（x，y）=h（x），（y）i，对于所有x，y∈x。

对于所有f0 H0和x x，f，κXi= f（x），

被称为复制财产的财产。

证据。对于任意两个线性组合，用x j，yk∈x和αj，βk∈c，定义hf，gi

.（？）

乍一看，上述表达式似乎依赖于f和g的线性表达式。

组合，但由于κ（xj，yk）=κ（yk，xj），观察

，（）

由于第一项和第三项对于表示f和g的所有线性组合都是相等的，因此我们得出结论（†）仅取决于f和g，而不取决于它们作为线性组合的表示。

很明显（†）定义了一种赫米特亮片形式。对于每一个f∈h0，我们有

，

因为κ是一个正定核。对于任何有限子集f1，…，h0的fn和任何z∈cn，

，

结果表明，从h0×h0到c的map（f，g）7→hf，gi是一个正定核。注意，对于所有f∈h0和所有x∈x，（†）意味着

，

被称为复制财产的财产。上面的意思是

Hκx，κyi=κ（x，y）。（）

通过对正定核（f，g）7→hf，gi的53.4号命题，我们得到

εHf，κXi 2，HF，FIHκX，κXI，

也就是说，

| f（x）2≤hf，fiκ（x，x）、

sinceso h f，f\_（ix=0）=意味着κx，我们有f（x）=0，——因为alli是一个Hermitian内积onx∈x，这意味着h−，h−0i，by（定义为（）和†）是正定的。因此，H−

κ（x，y）=h（x），（y）i，对于所有x，y∈x。

设h为完成h0的希尔伯特空间，使h0在h中稠密。图η：h→cx由

ε（f）（x）＝Hf，κXi

显然是线性的，它是内射的，因为家庭（κx）x x x跨越H中稠密的H0，因此它在H中也是稠密的，因此如果Hf，κXi＝0，对于所有x x，则F＝0。

如果我们用函数η（f）来识别函数f∈h，那么我们就有了再生性质。

HF，κX= F（x），对于所有的fh h和x x x都是。

如果x是有限的，那么cx是有限维。如果x是一个可分离的拓扑空间，如果k是连续的，那么可以证明h是一个可分离的希尔伯特空间。

另外，如果k:x×x→r是实对称正定核，那么我们马上就能看到定理53.8适用于h0的实欧几里德空间和h的实希尔伯特空间。

53.2。正核的希尔伯特空间表示

注：如果x=g，其中g是局部紧群，那么函数p:g→c（不一定连续）是正半定的，如果对于所有s1，…，sn∈g和所有ξ1，…，ξn∈c，我们得到

.

因此，如果我们用

κ（s，t）=p（t-1s），

那么k是g上的一个正定核。如果p是连续的，那么我们就知道p是由希尔伯特空间h中G群的一元表示u:g→u（h）产生的，其内积为h−、−i（具有一定连续性的同态），从某种意义上说，这里有一些向量x0。使

p（s）=hu（s）（x0），x0i，对于所有s∈g。

由于美国是H上的单一运营商，

p（t−1s）=hu（t−1s）（x0），x0i=hu（t−1）（u（s）（x0）），x0i

=hu（t）（u（s）（x0）），x0 i=hu（s）（x0）），u（t）（x0）i，

这表明

κ（s，t）=hu（s）（x0）），u（t）（x0）i，

因此，图\_：g→h由\_（s）=u（s）（x0）给出。

是特征空间H中的一个特征映射。这个定理是由GelfandRaikov（1943）提出的。

定理53.8的证明与Godement关于正型函数与一元表示之间对应关系的上述结果的部分证明基本相同；见Helgason[89]，第四章，定理1.5。定理53.8更为一般，因为它不假定x是一个群，但当g是一个群时，特征映射是由一个统一表示产生的。

集合上的内核可以用度量来定义。

例53.7。设（d，a）为可测量空间，其中d为非空集，a为d上的σ-代数（可测量集）。设x为a的一个子集。如果μ是（d，a）上的正测量值，并且μ是有限的，这意味着μ（d）是有限的，那么我们可以定义图k 1:x×x→r，由

κ1（a1，a2）＝（a1 a2），a1，a2 x。

我们可以证明κ是一个核函数，如下所示。设h=l2μ（d，a，r）为可积函数的希尔伯特空间，其内积为

Z

H

f，g=f（s）g（s）d礹（s），

D

并设\_：x→h为特征嵌入，由

⑨（a）=χa，a∈x，

A的特征函数，那么我们有

Z

κ1（a1，a2）＝（a1a2）＝χa1 a2（s）dµ（s）

D

Z

=χa1（s）χa2（s）dµ（s）=hχa1，χa2i

D

=h\_（a1），\_（a2）i.

上面的内核称为交集内核。如果我们假设μ是标准化的，那么μ（d）=1，那么我们也有联合补码内核：

κ2（a1，a2）=（a1 a2）=1（a1 a2）。

核κ1和κ2的总和κ3是一致核：

κS（a1，a2）=1−（a1−a2）−（a2−a1）。

可以设计许多其他类型的内核，特别是图形内核。有关内核的全面介绍，请参阅sch–olkopf和smola[141]以及shawe–taylor和Christianini[154]。

## 53.3核PCA

作为核函数的一个应用，我们讨论了主成分分析方法（PCA）的推广。假设我们在一些输入空间x中有一组数据s=x1，…，xn，并假设我们在（真实）特征空间（f，h−，−i）中嵌入了x的：x→f，但我们只能访问内核函数κ（x，y）=h（x），（y）i。我们想对集合（s）=（x）进行PCA分析。1），…，\_（xn）。

有两个障碍：

1. 我们需要将数据中心化，并计算成对中心数据的内积。更准确地说，如果\_（s）的质心是

，

然后我们需要计算内积h\_（x）−\_（y）−i。

53.3。核PCA

1. 让我们假设F＝RD与标准欧几里得内积，并且数据点（XI）表示为N×D矩阵X的行向量Xi（因为它是Curm）。内积κ（Xi，XJ）＝h（Xi），（xJ）i由核矩阵k= xxx给出。请注意，用这种表示，（Xi）是D维列向量，并且（Xi）＝Xi＞。然而，主分量yk（视为n维列向量）的jth分量（yk）j是由xbj=xj−μ投影到方向uk（视为d维行向量）给出的，该方向是矩阵的单位特征向量（x−μ）>（x−）（其中xb=x−）。礹是第j行为）的矩阵，由内积给出。

hxj−μ，uki=（yk）j；

见定义21.2（第一卷）和定理21.11（第一卷）。问题是，我们知道矩阵（x−礿）（x−礿）>是从（1）得到的，因为它可以用k表示，但我们不知道（x−礿）>（x−礿）是什么，因为我们没有访问xb=x−礿的权限。

这两个困难都很容易克服。对于（1），我们有

.

对于（2），如果K是核矩阵k=（κ（Xi，xJ）），则对应于核函数κ的核矩阵KB由

Bκ（x，y）=h（x）−\_，（y）−\_i b

可以用k表示。设1为列向量（维度n），其条目均为1。则11>是n×n矩阵，其条目均为1。如果a是n×n矩阵，那么1>a是由a列和组成的行向量，a1是由a列和组成的列向量，1>a1是a中所有项的和，那么很容易看出核函数k对应的核矩阵是给定的。通过

乙

K11。

假设xb=x−礹具有秩r。要克服第二个问题，请注意，如果

xb=v du>

是XB的SVD，那么

xb>=ud>v>

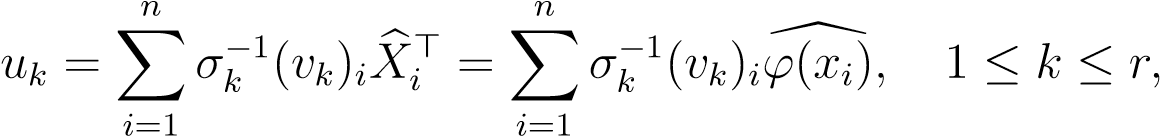
是xb>的一个svd，d>的r×r次矩阵由d>（和d）的第一行r和r列组成，是由xb的奇异值σ1≥·············································由V（1≤K≤R）的前r列v k组成

ur=xb>vr∑−r 1。

此外，是的非零特征值，并且vr的列是kb的相应的单位特征向量。从

ur=xb>vr∑−r 1

Ur的第k列Uk（是与特征值σk2相关的xb>xb的单位特征向量）由下式给出：



因此，根据下式给出了[（x）到英国的投影

\*n+

H[[（x），uCi]＝（x），x＝k＝1（vk）i〔（Xi）〕

i＝1

n

= x，k，k，1，（d），[（x）e（x）e＝x，k＝1，（i）。

I=1 I=1

因此，主分量YK在主方向UK的第j个分量由下式给出：

.

将核PCA推广到（实）特征空间（f，h−−i）中x的一般嵌入，其中核矩阵k由

Kij= H\*（Xi），（XJ）I，

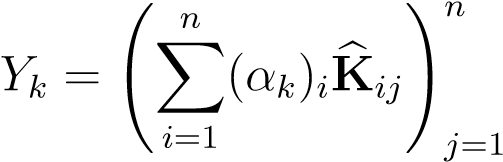
步骤如下。设r为kb的秩，其中

K11

设为kb的非零特征值，v1，…，vr为相应的单位特征向量。符号

αk=σk−1vk

通常使用，其中αk被称为双变量。列向量yk（1≤k≤r）由



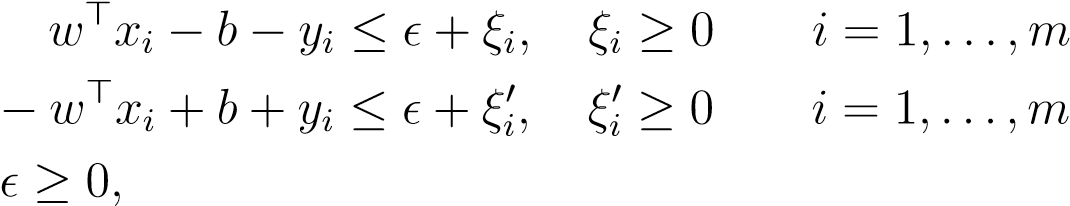
被称为数据集s=x1，…，xn在方向b（即使矩阵xb未知）的kth内核主成分（简称kth内核PCA）。

在下一节中，我们给出了在岭回归泛化中使用内核函数的另一个说明（参见第52.1节）。

## 53.4ν-sv回归

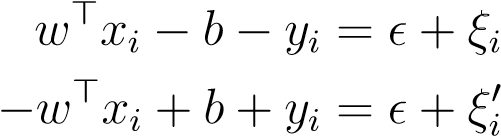
设{（x1，y1），…，（xm，ym）}是一组观测数据，通常称为训练数据集，用Xi→Rn和Yi-r r。我们的目标是学习F（x）＝w＞x－b的仿射函数F，它符合训练数据集，但不惩罚低于给定的0的误差。因此，我们尝试将一个具有半径的管与数据相匹配，但是我们也允许误差，在某种意义上，一些数据XI可以满足某些Zi＞0的相等，或者对于某些Zi0＞0的等式（f（Xi））。在这种情况下，XI位于管的半径之外。松弛变量ξi和ξi0的大小和大小之间的权衡是通过使用两个常数ν≥0和c>0来实现的。对于短的nv-sv回归，nv-支持向量回归的方法由以下最小化问题规定：nv-sv回归：

最小化



最小化变量w、b、ξ和ξ0。约束是仿射的。

首先，观察方程



只有在以下情况下才能同时保持

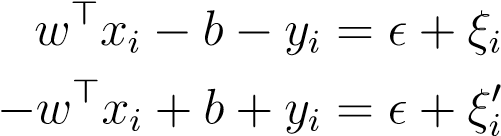
，

也就是说，

，

因为0，只有当=0时才会发生这种情况，然后

西> B＝易。特别是，如果>0，则方程式



不能同时保持。此外，由于w＞X+B+Yi＝（w＞Xi，b，y），对于最优解，如果w＞Xi→B→Yi为0，则Zi0 i0＝0，因为不等式成立。



基本满足（因为0），如果W\*Xi，B，Yi小于0，则类似地，Zi＝0。因此，我们得到了方程

（ξξ0）

观察如果v>1，则上述程序的最佳解必须产生

实际上，如果大于0，我们可以减少一小部分δ>0，并增加以满足约束条件，但目标函数的变化量为−vδ+δ，这是负的，因为0不是最佳值。

驱动到零不是目标，因为通常数据不是噪声的，所以非常少的对（Xi，Yi）满足方程W＞X-B＝Yi，然后许多对（Xi，Yi）将对应于一个误差（Zi i＞0或ZiI0＞0）。因此，我们通常假设0<ν≤1。

为了构造拉格朗日，我们将拉格朗日乘子αi＞0分配到约束W\*Xi，拉格朗日乘子αi0±0以上的约束条件，

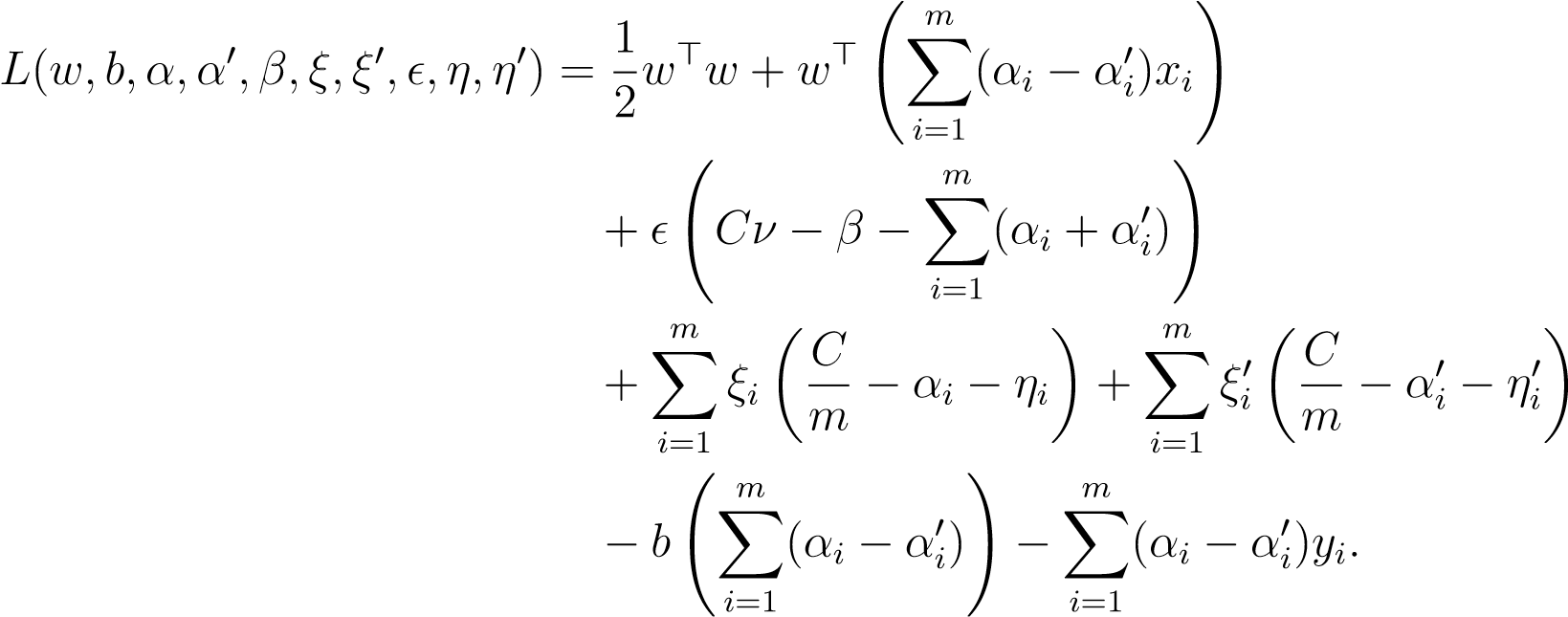
拉格朗日乘数ηi≥0到约束ξi≥0，拉格朗日乘数

约束0，拉格朗日乘子β≥0到约束0。这个

拉格朗日是

，

拉格朗日也可以写成



为了找到关于原始变量的双函数g（α，α0，η，η0，β），我们将其最小化。观察拉格朗日是凸的，由于（，一个凸的开集，根据定理39.11，拉格朗日有一个最小的iff=0，所以我们计算梯度。我们得到

，

哪里

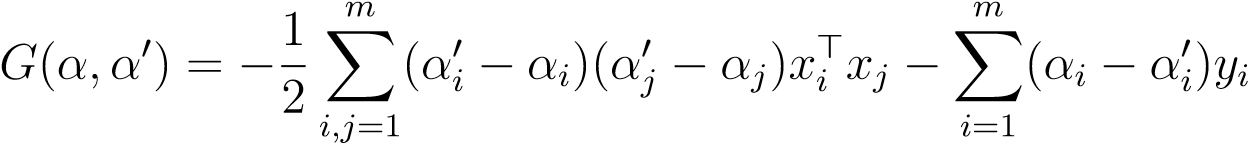
，和。

因此，如果我们设为0，我们得到方程

，（W）

.

将上述方程代入拉格朗日方程的第二个表达式中，我们发现双函数g与变量β、η、η0无关，由下式得出：



如果

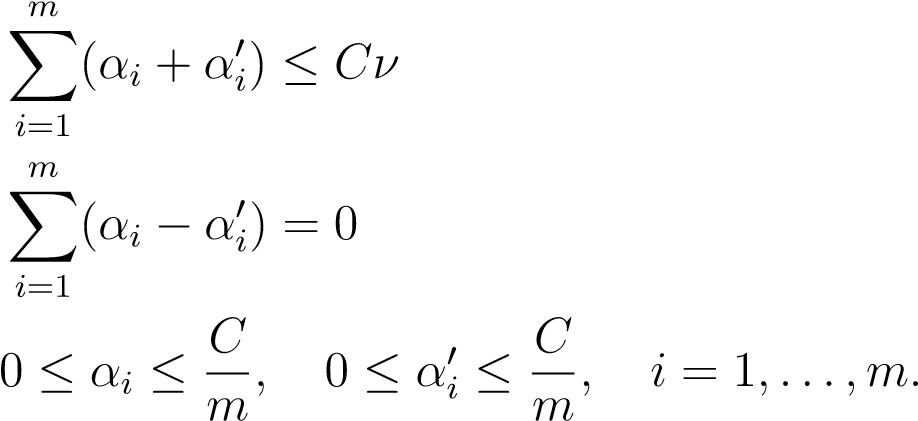
，

以及−∞否则。

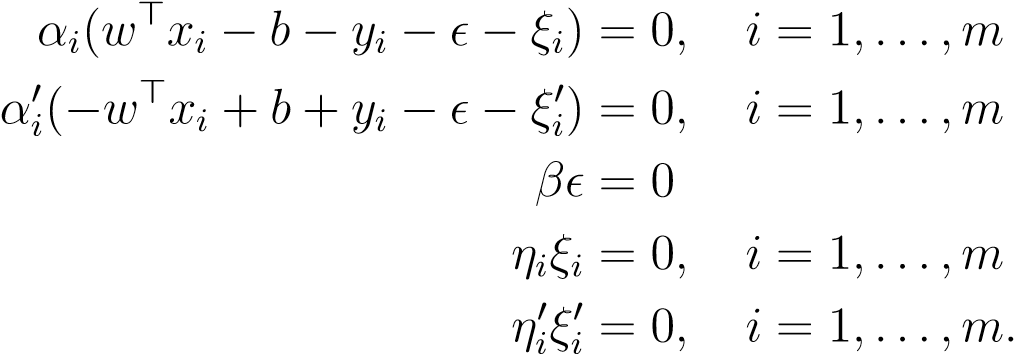
对偶程序是通过使g（α，α0）最大化或等效地通过最小化得到的。

-以下双程序：G（α，α0），结束。考虑到η，η0≥0和β≥0，我们得出

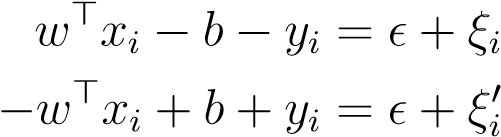
最小化



KKT条件（针对主要项目）是



如果>0，因为方程



不能同时保持，我们必须

（αα0）

从方程式中

，

我们得到方程

（三）

这些方程表明，如果ξi>0，那么，我们就有了活动约束。



并且XI是一个错误，类似地，如果ZEII0> 0，那么，我们就有活动约束。

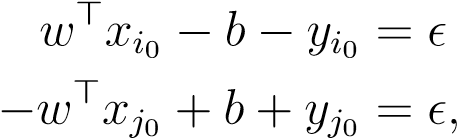


XI是一个错误。

如果原始解具有w=06且>0的最优解，则通过（w）和自

）=0和αiαi0=0，

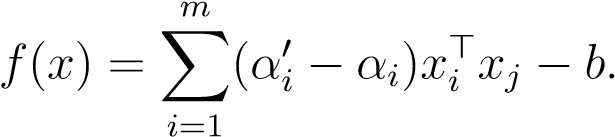
有一些i0，比如αi0>0，一些j0=6i0，比如0。在温和假设下，有一些i0，比如0，有一些j0，那么通过（），我们得到了ξi0=0，ξj0 0=0，我们得到了这两个方程。



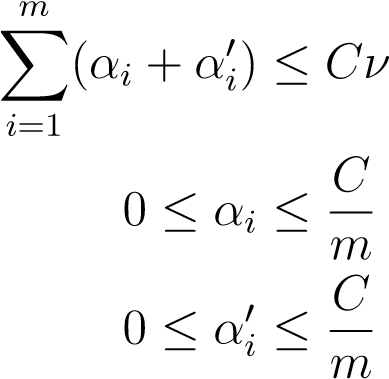
所以可以计算b和。特别地，

.

函数f（x）=w>x−b（通常称为回归估计）由以下公式给出：

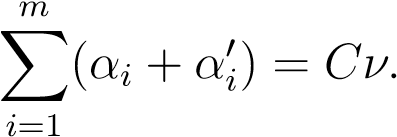


制约因素



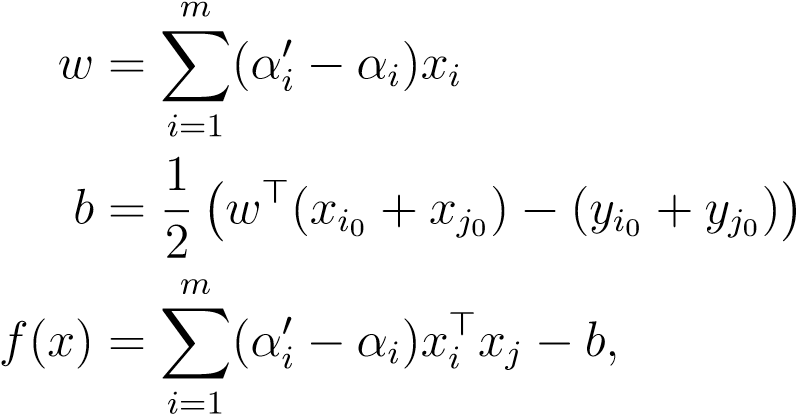
这意味着数据中至多有一个分数。如果随后的结果是，如果>0和0<ν≤1，那么ν是误差分数的上界。

kkt条件意味着如果>0，那么β=0，在这种情况下



因为=0，并且由于支持向量对应于0，我们看到，ν是支持向量分数的下界。

因为w、b和f（x）的公式，

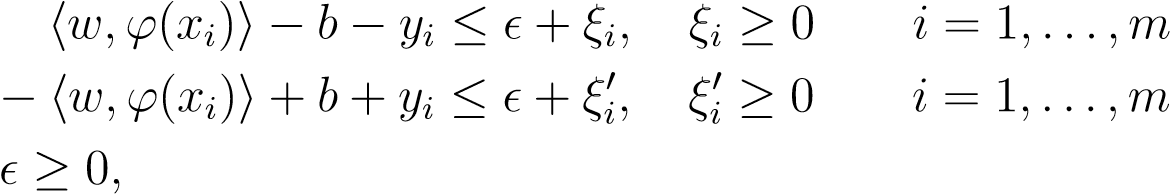


仅在数据点XI中涉及内积，并且由于对偶程序的目标函数g（α，α0）也只涉及数据点XI中的内积，所以我们可以对其进行回归。

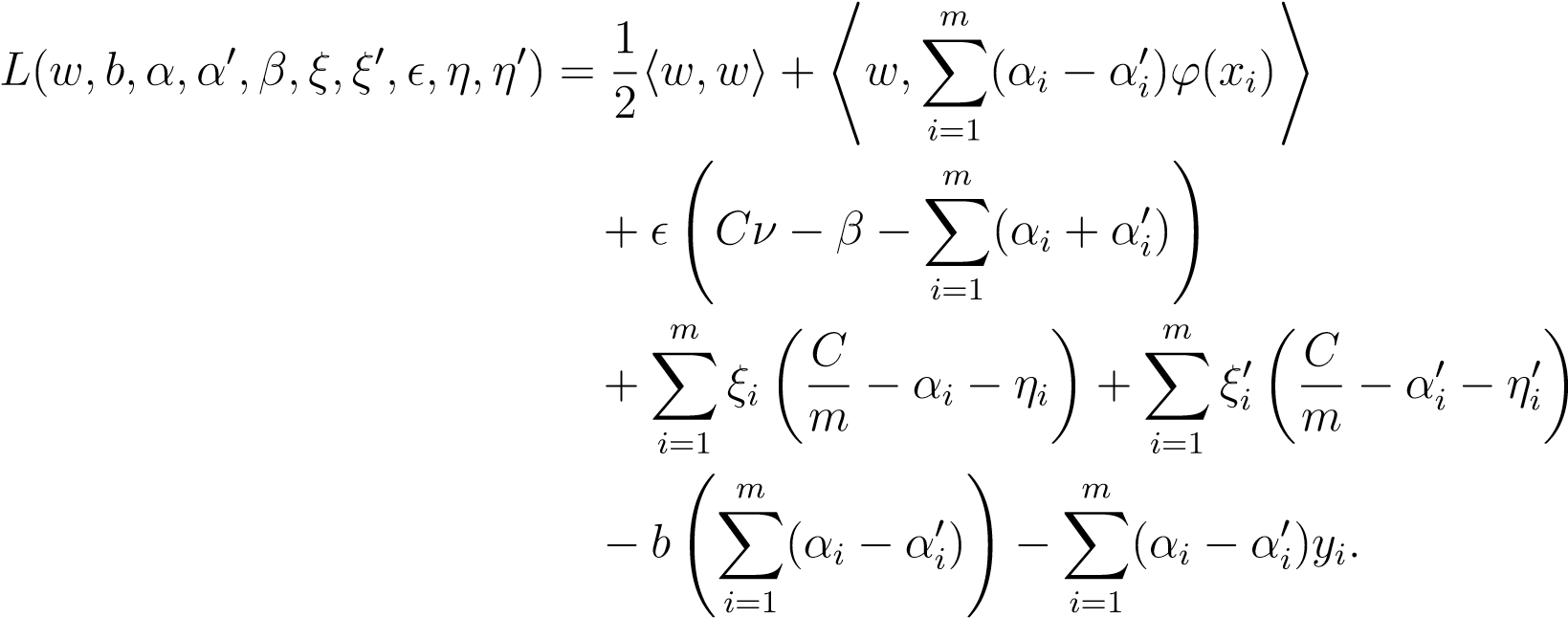
与前面的部分一样，我们假设我们的数据点{x1，…，xM}属于一个集合x，我们假设我们具有特征空间（f，h，，i）和特征嵌入映射：x，f，但是我们只能访问核函数κ（Xi，xJ）＝h（x），（xJ）i。n数据集上的特征空间f（（x1），y1），…，（（xm），ym）。经过前面的计算，我们看到原始程序是由

核v-sv回归：

最小化



最小化变量w、b、ξ和ξ0。拉格朗日由



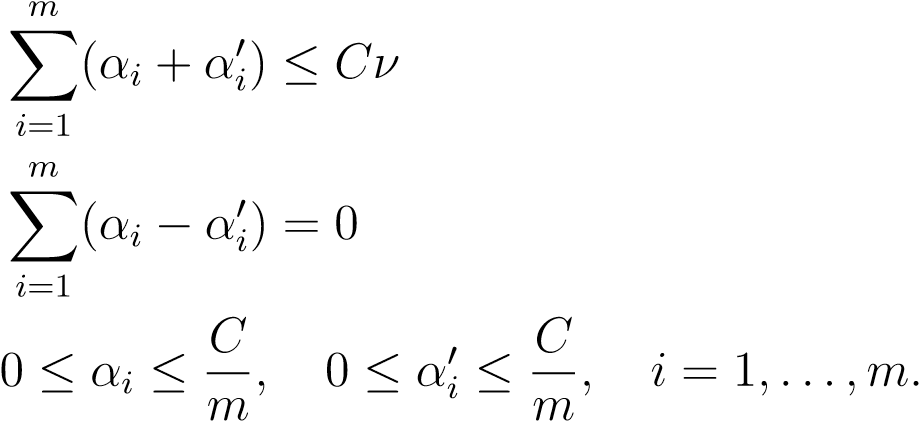
将拉格朗日梯度设为零，也得到了方程组。

，（W）

.

利用上述方程，我们发现双函数g与变量β、η、η0无关，得到以下双程序：

最小化



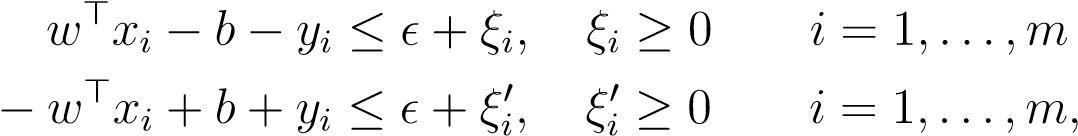
我们前面所说的一切也适用于核Vo-SV回归方法，除了Xi被（Xi）代替，并且必须使用内积h，，i，并且我们有公式。

！

只涉及κ的表达。

注：通过设置ν=0并保持不变，得到了一个关于ν-sv回归的变量。这种方法称为-SV回归或（线性）不敏感SV回归。相应的优化程序为-SV回归：

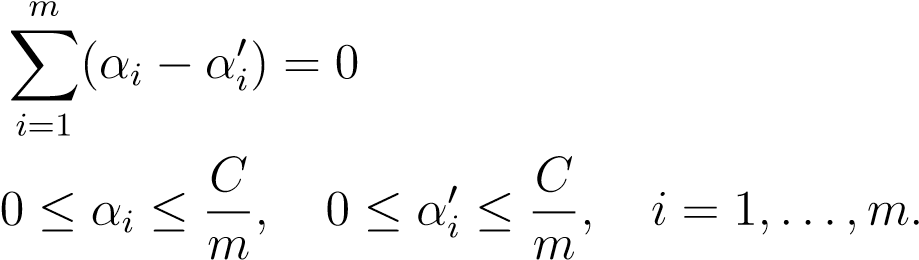
最小化



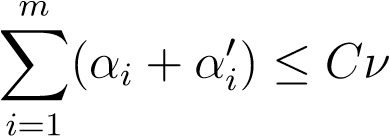
最小化变量w、b、ξ和ξ0。

很容易看出双程序是

最小化



约束条件



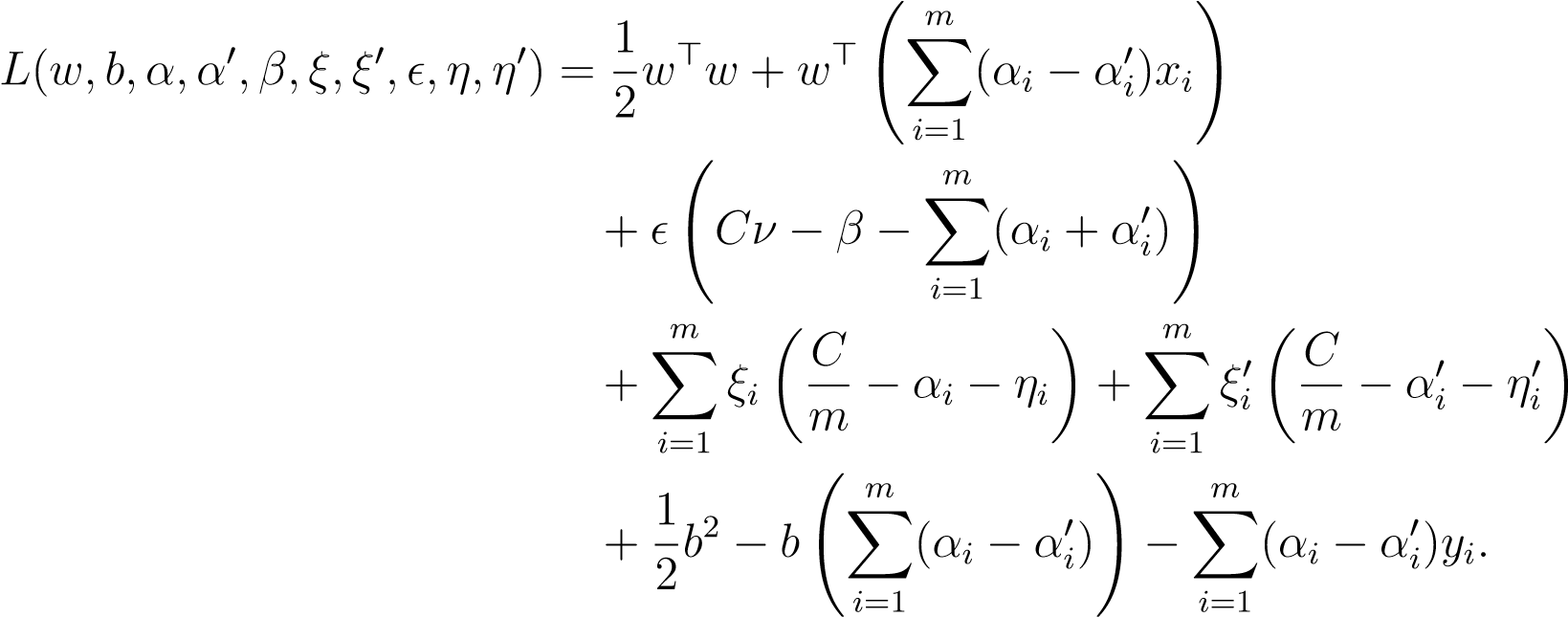
已经不存在了，但是额外的项）已经被添加到双重功能中，以防止αi和爆炸。

有一个明显的-sv回归的核心版本。可以很容易地看出，如果ν-sv回归成功并得到w，b，>0，那么具有相同c和相同值的-sv回归也成功并返回相同对（w，b）。有关这些方法的更多详细信息，请参阅sch–olkopf、smola、williamson和bartlett[143]。

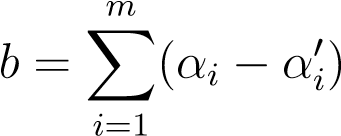
注：线性惩罚函数）可以用二次惩罚函数来表示；见Shawe–Taylor和Christianini[154]（第7章）。

然而，v-sv回归的另一个变体是将该项添加到目标函数中。

新拉格朗日是



我们得到了新的方程

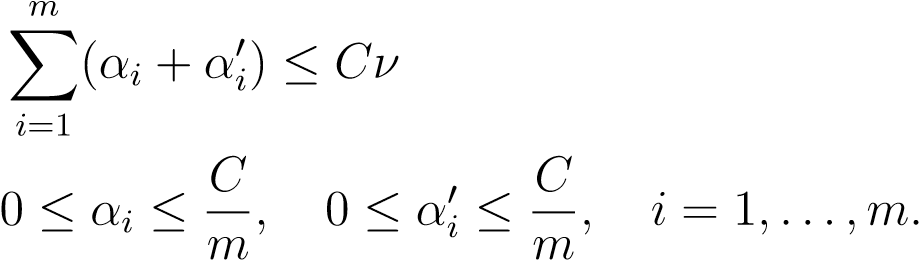


确定b，替换方程式

.

新的双重计划是

最小化



第五十四章

# 软边界支持向量机

如果点集u1，…，up和v1，…，vq不是线性可分的（使用ui，vj∈rn），我们可以使用线性规划的一个技巧，即引入非负的“松弛变量”来放松“硬”约束。

w>ui−b≥δi=1，…，p

−w>vj+b≥δj=1，…，q

从第49.5节到“软”约束的问题（SVMH1）

回想一下

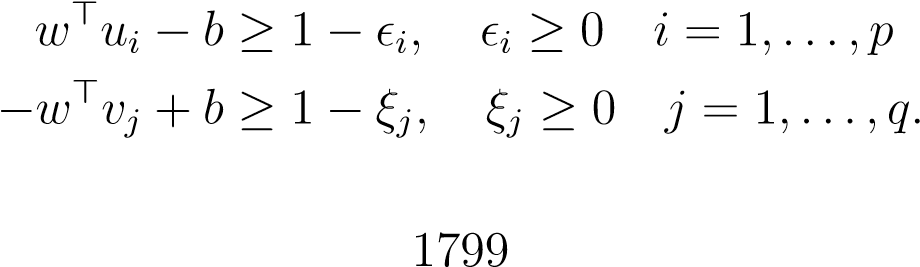
如果为0，则点UI可能被错误分类，从某种意义上说，它可能属于边缘（板），甚至属于对负（红色）点进行分类的错误半空间。见图54.1（2）和（3）。同样，如果ξj>0，点vj可能被错误分类，从某种意义上说，它可能属于边缘（板），甚至属于对正（蓝）点进行分类的错误半空间。我们可以将i视为违反约束w>ui−b≥δ的程度的度量，而将ξj视为违反约束w>vj+b≥δ的程度的度量。如果=0，则恢复原始约束。通过足够大，这些约束总是可以满足的。我们加上约束w>w≤1，然后最小化−δ。

如果不是问题的约束（svmh1），我们使用硬约束

w>ui−b≥1 i=1，…，p

−w>vj+b≥1 j=1，…，q

问题（svmh2）（参见示例49.6），然后我们放松到软约束



在这种情况下，对w没有约束，但我们将（1/2）w>w最小化。

理想情况下，我们希望找到一个分离超平面，以最小化错误分类点的数量，这意味着变量应尽可能小，但在最大化利润（板的厚度）和最小化错误分类点的数量上存在权衡。这反映在目标函数的选择上，根据我们是最小化变量的线性函数，还是这些变量的二次函数，或者是否在目标函数中包含术语（1/2）b2，有几个选项。这些方法被称为支持向量分类算法（简称SVC算法）。

SVC算法寻求一个“最优”分离方程w>x-b=0的超平面h。如果有新的数据x∈rn出现，我们可以通过计算量w>x−b的符号，确定由超平面h确定的两个半空间中属于哪一个来对其进行分类。函数sgn:r→−1,1由下式给出：

.

然后我们定义了与方程w>x−b=0的超平面h相关的（二进制）分类函数，即f（x）=sgn（w>x−b）。

值得注意的是，所有已知的寻找超平面的优化问题都具有这样的性质：权重向量w和常数b是由只涉及输入数据点ui和vj内积的表达式给出的，分类函数也是如此

f（x）=sgn（w>x-b）。

这是一个关键的事实，它允许使用内核方法对支持向量机进行广泛的推广。

内核的方法包括假设输入空间rn嵌入一个更大（可能是无限维）的欧几里得空间f（内积h−−−i），通常称为特征空间，使用函数

⑨：RN→F

称为功能图。函数k:rn×rn→r由

κ（x，y）=h（x），（y）i

是嵌入的核心函数，见第53章。我们的想法是，特征映射“展开”输入数据，使其在高维空间f中更具线性。现在，即使我们不知道特征空间f是什么以及嵌入映射是什么，我们也可以假装在f中解决嵌入数据点的分离问题。ts\_（ui）和\_（vj）。因此我们求方程的超平面H

hw，ζi−b=0，ζ∈f，

在特征空间F中，尝试分离点（ui）和点（vj）。如我们所说，结果表明，w和b是通过只涉及内积的表达式给出的，其中κ（ui，uj）=h（ui），（uj）i，κ（ui，vj）=h（ui），（vj）i，和κ（vi，vj）=h（vi），（vj）i形成对称（p+q）×（p+q）矩阵k（核矩阵），由下式给出

）1≤I≤P，1≤J≤Q

K

κ（vi−p，vj−q）p+1≤i≤p+q，p+1≤j≤p+q。

然后是分类功能

f（x）=sgn（hw，（x）i-b）

对于原始数据空间中的点，Rn也仅用矩阵k和内积κ（ui，x）=h（ui），（x）i和κ（vj，x）=h（vj），（x）i表示。因此，在原始数据空间Rn中，超曲面

S=X∈RN HW，（X）I−B=0

分离数据点ui和vj，但它不是rn的仿射子空间。分类函数f告诉我们S的“边”是一个新的数据点x∈rn。因此，我们设法通过假设存在一个嵌入的瓒：rn→f，即使我们不知道它是什么，但通过给定的核函数κ：rn×rn→r可以访问f，将不可被仿射超平面、非仿射超曲面分离的数据点ui和vj分开。由内积k（x，y）=h（x），（y）i。

在实践中，使用内核方法的艺术就是选择正确的内核（就像《印第安纳琼斯》中的骑士所说的那样，“明智地选择”。

内核的方法非常灵活。它还适用于支持向量机的软边缘版本，也适用于回归问题、主成分分析（PCA）以及机器学习中出现的其他问题。

在sch–olkopf和smola[141]以及Shawe–Taylor和Christianini[154]中发现了对内核方法的全面介绍。另见Bishop[23]。

首先考虑问题产生的软边际支持向量机（SVMH1）。

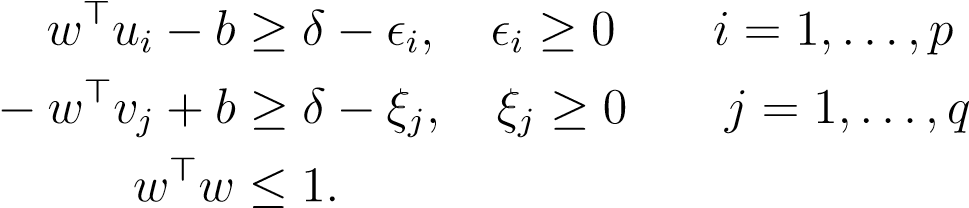
## 54.1软边界支持向量机；（SVMS1）

在本节中，我们推导了与以下版本的问题（svmh1）产生的软边值svm相关的双函数g，其中边值δ的最大化被−δ的最小化所取代，并且在这里我们添加了一个“正则化项”，其目的是使和ξ∈RQ稀疏（即，尝试

make有尽可能多的零），其中k>0是一个固定常数，可以调整以确定这个正则化项的影响。如果原问题（svms1）有一个最优解（w，δ，b，，ξ），我们试图用对偶函数g来获得它，但是我们可以看到，在这个问题的特殊公式中，约束w>w≤1会引起问题，即使它是凸的。

软利润SVM（SVMS1）：

最小化



习惯上写`=p+q。

对于这个问题，原问题可能有一个Kwk=1且δ>0的最优解（w，δ，b，，ξ），但如果点集不是线性可分的，则对偶的最优解可能不会产生w。

我们问题的目标函数是仿射，唯一的非仿射约束w>w≤1是凸的。这个约束是合格的，因为对于任何w=06，这样w>w<1，对于任何δ>0和任何b，我们可以选取足够大的，从而满足常数。因此，根据定理49.16（2），如果原始问题（SVMS1）有一个最优解，那么对偶问题也有一个解，对偶间隙为零。

不幸的是，这并不意味着对偶的一个最优解产生一个原始解的最优解，因为定理49.16（1）的假设不成立。一般来说，可能没有一个唯一的向量（w，，ξ，b，δ），使得

.

如果集合ui和vj不是线性可分的，那么对偶问题可能有一个解γ=0，

，

和

，

因此定义了双函数g（λ，μ，α，β，γ），它是一个偏函数，其值为g（λ，μ，α，β，0）=0。这样一对（λ，μ）对应于两个凸组合的系数

PQ

十倍

2λiui=2μjvj

I=1 J=1

它对应于凸壳conv（u1，…，up）和conv（v1，…，vq）的（非空）交叉点中的相同点。结果表明，W与对偶函数的唯一联系是方程

，

当γ=0时，这个方程为0=0，所以对偶问题对于确定w是无用的。这一点在文献中似乎被遗漏了（例如，在Shawe–Taylor和Christianini[154]第7.2节）。对偶问题证明了δ≥0。然而，如果γ=06，则w由双组分的任何溶液（λ，μ）测定。

仍然需要计算δ和b，这可以在我们称之为标准裕度假设的温和假设下进行。

如果（w，δ，b，，ξ）是问题（svms1）的最优解，则将点ui和vj分为以下几类：

=0，则点ui被正确分类，位于蓝色边缘（方程w>x=b+η的超平面hw，b+η）或蓝色边缘的正确侧（蓝色侧）。同样，如果ξj=0，则点vj被正确分类，并位于红边（方程w>x=b−η的超平面hw，b−η）或红边（红边）的正确一侧。

（2）如果为0，则点ui位于边缘（板）内，但位于分离超平面的正确侧（蓝色侧）。如果，那么UI位于分离的超平面上。同样，如果0<ξj≤η，则点vj位于边缘（板）内，但位于分离超平面的正确侧（红色侧）。如果ξj=η，则vj位于分离超平面上。

，则点ui位于分离超平面的错误一侧（红色一侧）；它被错误分类。同样，如果ξj>η，那么点vj位于分离超平面（蓝色面）的错误一侧；它被错误分类。

设为与不等式相关联的拉格朗日乘数，设为拉格朗日乘数与不等式相关联−w>vj+b≥δ−ξj，设为与不等式相关联的拉格朗日乘数为与不等式相关联的拉格朗日乘数ξj≥0，a让γ∈r+是与不等式w>w≤1相关的拉格朗日乘子。

线性约束由2（p+q）×n+p+q+2）矩阵给出，该矩阵以块形式给出：

，

其中x是n×（p+q）矩阵

，

线性约束表示为

.

更明确地说，c是以下矩阵：

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| −U>1  .  …  γ  −U>P  γ  v1>  γ  ……  γ  γ | 1  …  零  零  … | ···…  ·········  ···… | 零  …  1  零  … | 零  …  0比1  … | ···…  ·········  ···… | 零  …  零  零  … | 一  …  1比1  … | 1℃  \_  1\_  1\_  \_  γ |

VQ>0·····0··1·····C=

0−1····0 0··0 0 0··

 

…………………………………………………………

γ

 

0 0························0 0 0

 

0 0·····0····1···0 0 0··

\_

0 0·············0 0 0

目标函数由

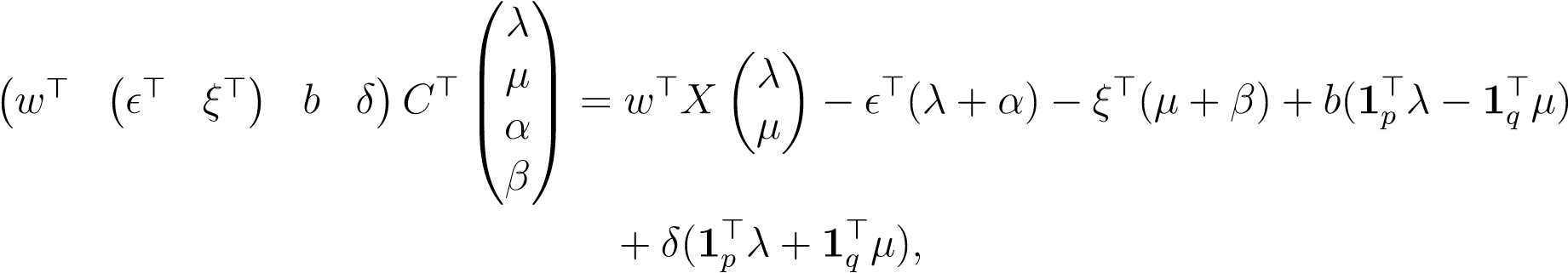
.

给出了，和γ∈r+的拉格朗日

通过

.

自从



拉格朗日可以写成

.

为了找到双函数g（λ，μ，α，β，γ），我们将w，ξ，b和δ最小化。因为拉格朗日是凸的（

r×r，一个凸开集，根据定理39.11，拉格朗日在（iff=0）中有一个最小值，因此我们计算了关于w，，ξ，b，δ的梯度，得到

.

通过设置=0，我们得到方程

（W）

和

λ+α=k1pμ+β=k1q

1>pλ=1>qμ1>pλ+1>qμ=1。

第二和第三个方程等价于不等式

0≤λi，μj≤k，i=1，…，p，j=1，…，q，

通常称为箱约束，第四和第五个方程产生

.

首先让我们考虑一下奇异情况γ=0。在这种情况下，（w）意味着

，

而术语γ（w>w−1）在拉格朗日中缺失，考虑到上面的其他四个方程，它减少到

.

总之，我们证明了如果γ=0，那么

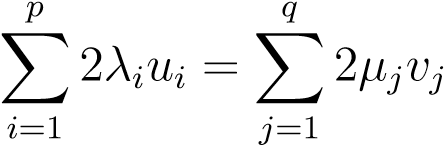
γ

\_\_\_\_0，如果g（λ，\_，α，β，0）=

\_\_\_\_−∞其他方式和。

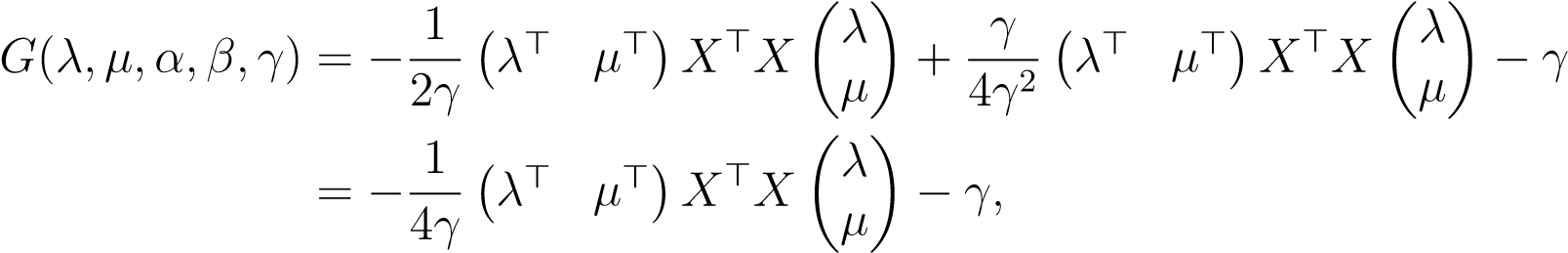


几何上，（λ，μ）对应于两个凸组合的系数

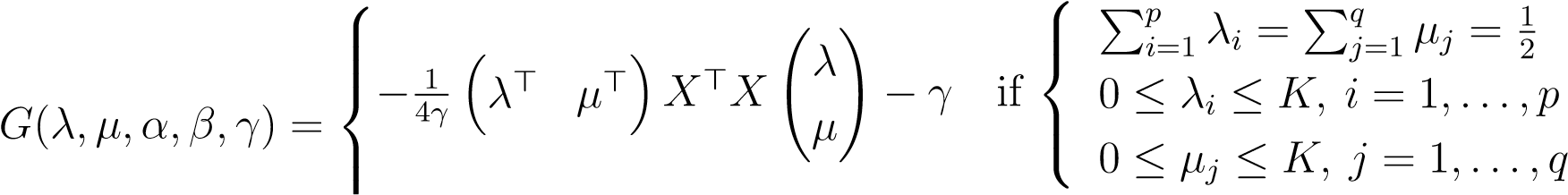


与凸壳conv（u1，…，up）和conv（v1，…，vq）交叉点的相同点相对应，如果集合ui和vj不是线性可分的。如果集合ui和vj是线性可分离的，那么凸壳conv（u1，…，up）和conv（v1，…，vq）是不相交的，这意味着γ>0。

现在假设γ>0。把方程（w）中的w插回到拉格朗日方程中，简化后，我们得到



因此，如果γ>0，双函数独立于α，β，并由下式得出：



−∞否则。

因为x>x是对称正定的，γ≥0，显然

g（λ，μ，α，β，γ）≤0

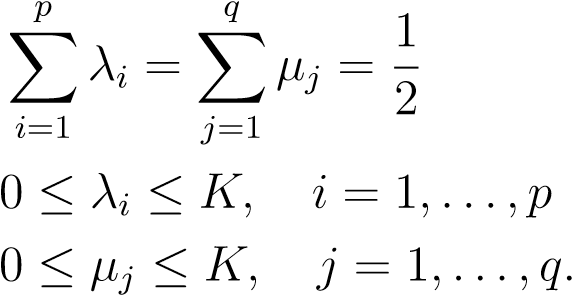
对于所有γ>0。

双程序由

最大化

γ=0时为0

从属于



同样，如果γ=0，那么

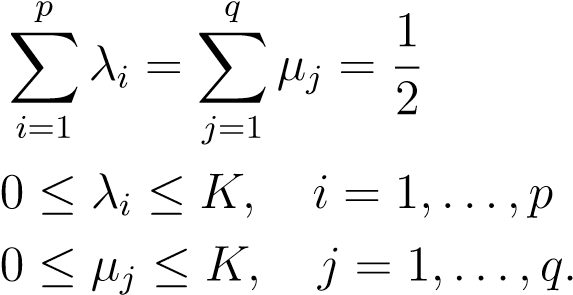
关于γ>0的最大化产生

所以我们得到

.

最后，由于g（λ，μ）=0和=0，双程序相当于以下最小化程序：

最小化

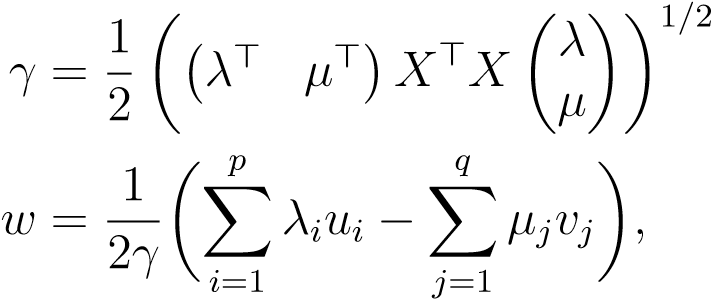


注意，约束意味着必须选择k，以便

.

通过使用基于梯度下降的数值程序（例如第51.6节中的ADMM）来求解双程序。如果原始问题是可解的，则会得到λ和μ的解。

如果最佳值为0，则γ=0和=0，因此在这种情况下不可能确定w。但是，如果最佳值大于0，则一旦通过（w）获得λ和μ的溶液，我们就可以



所以我们得到

，

这是生成单位向量的结果，因为

.

仍然需要找到b和δ，这不是由对偶程序给出的。

互补松弛条件根据λ和μ的值对点进行分类。事实上，我们有。

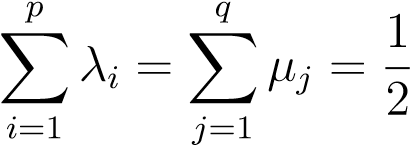
此外，如果λi>0，则相应的约束处于活动状态，同样，如果μj>0。因为

）=0，由于μj+βj=k，我们得到i，如果ξj>0，则μj=k。

因此，如果λi<k，则正确分类；同样，如果μj<k，则ξj=0和vj正确分类。我们有以下分类：

1. 如果0<λi<k，则ui在边缘，分类正确。同样，如果0<μj<k，则Vj在边缘，并且分类正确。
2. 如果λi=k，那么如果点ui可以正确分类，或者它位于正确边上的空白范围内，但是如果这样，它就被错误分类了。同样，如果μj=k，则如果ξj≤δ，则点vj可正确分类，或位于正确侧的边缘内，但如果ξj>δ，则其分类错误。
3. 如果λi=0，则ui被正确分类。同样，如果μj=0，则Vj分类正确。

方程式



意味着存在一些i0，例如λi0>0和一些j0，例如μj0>0，但是先验的，没有什么能阻止所有非零的λi的λi=k或所有非零的μj的μj=k的情况。如果发生这种情况，我们可以用更大的k值重新运行优化方法。如果下面的温和假设sis保持，然后可以找到b和δ。

（SVMS1）的标准裕度假设。有一些指数i0，如0<λi0<k，有一些指数j0，如0<μj0<k。这意味着一些ui0被正确分类，在蓝边上，一些vj0被正确分类，在红边上。

如果（svms1）的标准裕度假设成立，那么=0和祆j0=0，那么我们就有了活动方程。

w>ui0−b=δ和−w>vj0+b=δ，

我们得到b和δ的值

.

如前所述，定理49.16（2）的假设成立，所以如果原始问题（SVMS1）有一个w=06的最优解，那么对偶问题也有一个解，对偶间隙为零。因此，对于最优解，我们有

，

也就是说

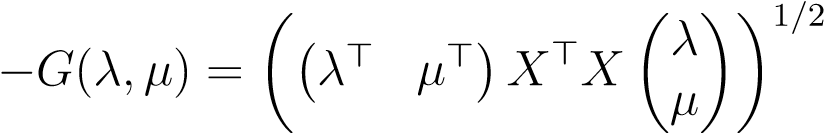
，

所以我们得到

.

因此，我们确定δ≥0。

需要注意的是，双重程序的目标功能



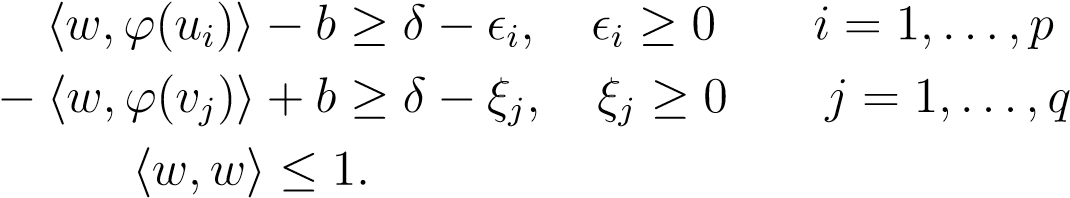
通过矩阵x>x只涉及到ui和vj的内积，同样，最优超平面方程也可以写成

，

一种只涉及x的内部产品与ui和vj以及ui和vj的内部产品的表达式。

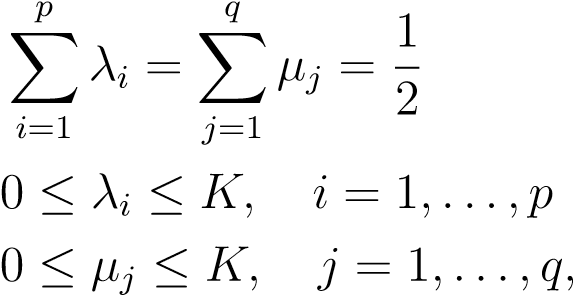
如本章开头所述，这是一个关键的事实，允许使用内核方法对支持向量机进行泛化。我们可以定义以下“内核化”版本的问题（SVMS1）：软边界内核SVM（SVMS1）：

最小化



通过计算，我们找到双程序的以下版本：

最小化



式中，k是`×`核对称矩阵（其中`=p+q），由

）1≤I≤P，1≤J≤Q

K

κ（vi−p，vj−q）p+1≤i≤p+q，p+1≤j≤p+q。

我们也发现

.

在标准裕度假设下，有一些指数i0，如0<λi0<k，有一些指数j0，如0<μj0<k，我们得出b和δ的值为

.

利用上述w值，我们得到

.

因此，分类功能

f（x）=sgn（hw，（x）i-b）

由给出

，

仅用核κ表达。

支持向量机的核心方法在sch–olkopf和smola[141]以及Shawe–Taylor和Christianini[154]中进行了讨论。

由于约束w>w≤1引起了麻烦，我们将其交换为另一个目标函数，其中−δ替换为（1）。这样我们就只剩下纯仿射约束。在下一节中，我们将讨论通过添加线性正则化项得到的问题（SVMH2）的推广。

## 54.2软边界支持向量机；（SVMS2）

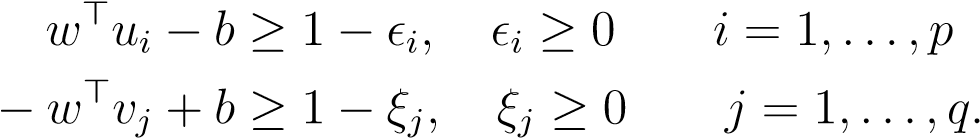
在这一节中，我们考虑问题的推广（svmh2），其中我们最小化

（1/2）w>w通过为一些k>0添加“正则化项”。

记住，裕度δ由δ=1/kwk给出。

软利润SVM（SVMS2）：

最小化



这是所有机器学习或模式分析书籍中讨论的经典问题，例如Vapnik[176]、Bishop[23]和Shawe–Taylor和Christianini[154]。所有变量都为0的平凡解被排除在外，因为不等式中存在1，但不清楚如果（w，b，，ξ）是最优解，那么w=0.6

证明了如果原问题有一个w=06的最优解（w，，ξ，b），则w由对偶的任何最优解（λ，μ）确定。我们还证明，对于一些i，λi>0，而对于一些j，μj>0。在我们称之为标准保证金假设的温和假设下，可以找到B。

如果（w，，ξ，b）是问题的最优解（svms2），那么点ui和vj的分类如下：

1. 如果=0，则点ui被正确分类，位于页边距或页边距的正确一侧（蓝色一侧）。同样，如果ξj=0，则点vj被正确分类，并且位于边缘或边缘的正确一侧（红色一侧）。见图54.1（1）。
2. 如果为0 1，则点UI位于边界（板）内，但位于分离超平面的正确一侧（蓝色一侧）。如果=1，则UI位于分离超平面上。同样地，如果0<ξj≤1，则点vj位于边缘（板）内，但位于分离超平面的正确侧（红色侧）。如果ξj=1，则vj位于分离超平面上。见图54.1（2）。

1，则点ui位于分离超平面（红色侧）的错误一侧；它被错误分类。同样，如果ξj>1，则点vj位于分离超平面（蓝色面）的错误一侧；它被错误分类。见图54.1（3）。

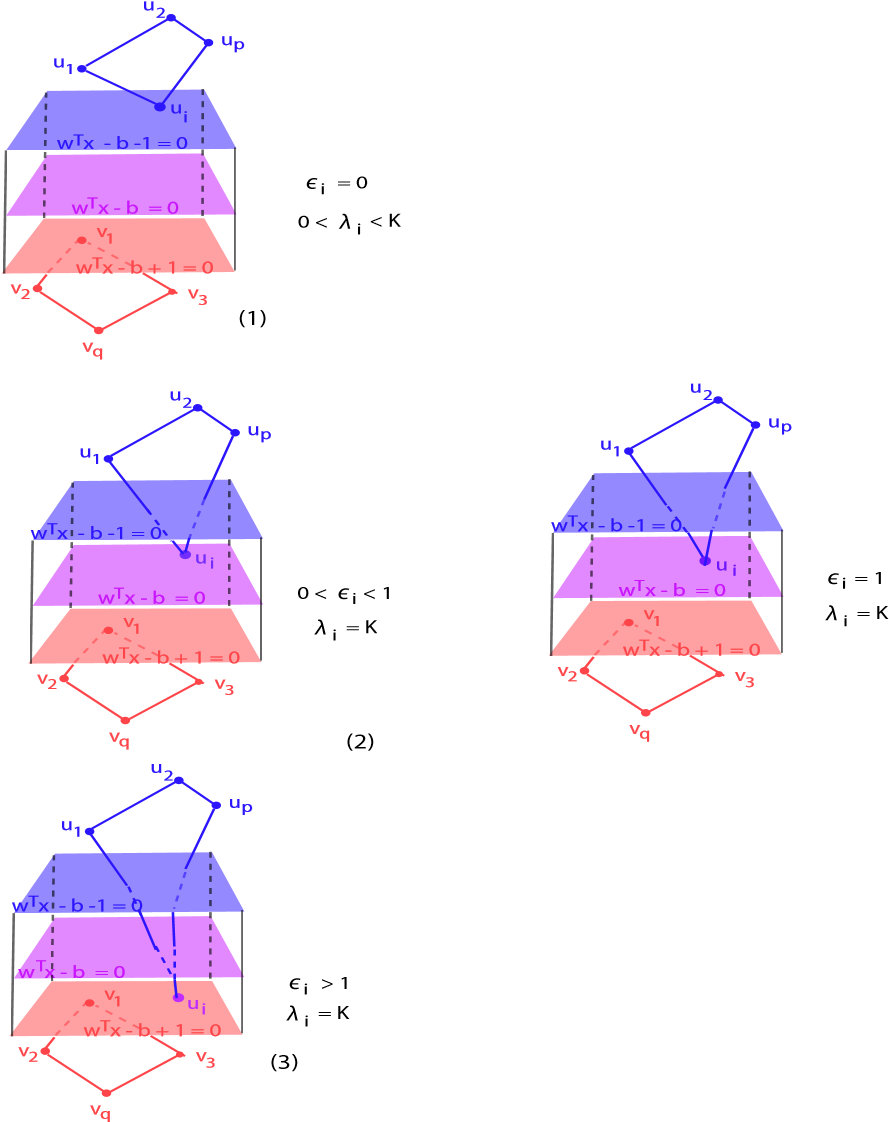


图54.1：图（1）说明了页边距中包含的UI的情况，并在

= 0。图（2）的左图是当ui在边界内，但仍在分离超平面w>x−b=0的正确一侧时；这在0 1时发生。右图描述了分隔超平面上的UI，只要=1。图（3）说明了用户界面的错误分类，并在

0）的点称为边缘误差；它们要么位于板内，要么被错误分类。

注意，这个框架对异常值仍然有些敏感，因为对错误分类的惩罚是线性的in和ξ。

首先，我们用矩阵形式写约束。2（p+q）×n+p+q+1）矩阵c以块形式写为

，

约束条件表示为

.

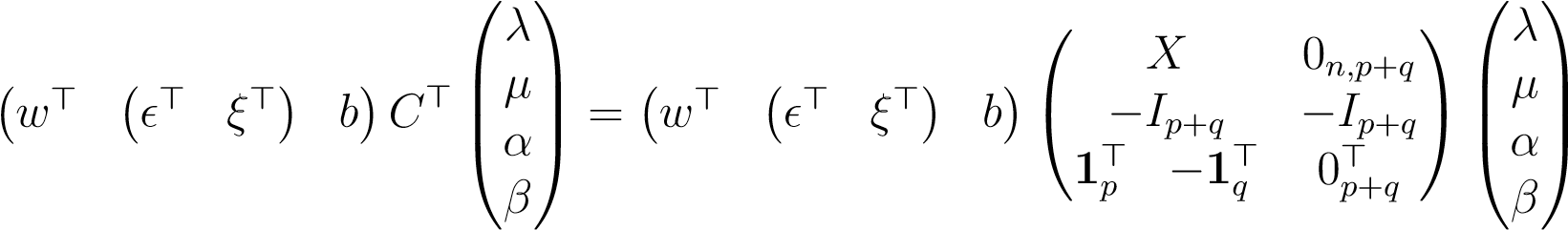
目标函数）由

.

带和带的拉格朗日由下式给出

.

自从



我们得到

，

从那以后

，

拉格朗日可以改写为



为了找到双函数g（λ，μ，α，β），我们将其最小化

. 由于拉格朗日是凸的（，一个凸的开集），根据定理39.11，拉格朗日在（=0）中有一个最小值，所以我们计算了它的梯度，得到

.

通过设置=0，我们得到方程

（W）

和

λ+α=k1pμ+β=k1q

1>pλ=1>qμ。

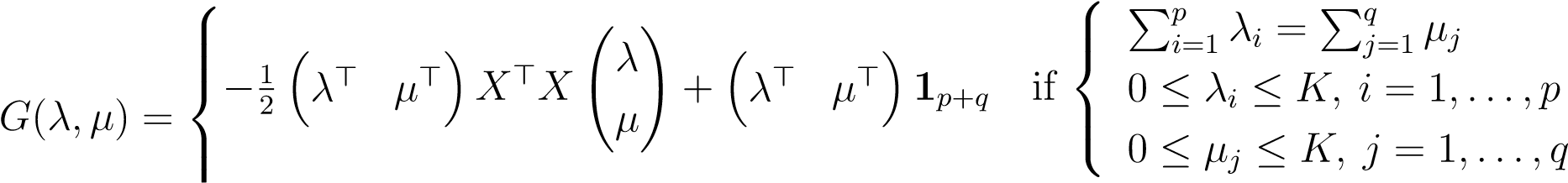
第一个和第四个方程与我们在示例49.10中得到的方程（1）和（2）相同。由于λ、μ、α、β≥0，第二个和第三个方程等于箱约束

0≤λi，μj≤k，i=1，…，p，j=1，…，q.

利用我们刚推导的方程，经过简化，我们得到

，

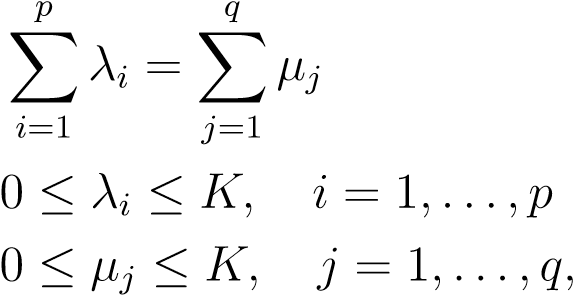
不依赖于α和β，与实施例49.10（4）中获得的双函数相同。要非常严格，



−∞否则。

如例49.10所示，双方案可表述为

最大化受试者



或同等

最小化

PQ

十倍

λi=μj i=1 j=1 0≤λi≤k，i=1，…，p

0≤μj≤k，j=1，…，q.

通过使用基于梯度下降的数值程序（例如第51.6节中的ADMM）来求解双程序。如果原始问题是可解的，则会得到λ和μ的解。

注：硬边值问题（svmh2）对应于特殊情况（svms2），其中=0，k=+∞。事实上，在问题（svmh2）中，拉格朗日方程中缺少涉及的项，其结果是缺少盒约束；我们只得到λi≥0和μj≥0。

我们可以用对偶程序来求解初等问题。一旦发现λ≥0，μ≥0，w由下式得出：

.

互补松弛条件根据λ和μ的值对点进行分类。事实上，对于j=1，…，q，我们有且ξjβj=0。

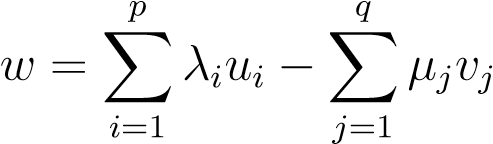
此外，如果λi>0，则相应的约束处于活动状态，同样，如果μj>0。因为

）=0，由于μj+βj=k，我们得到i，如果ξj>0，则μj=k。

因此，如果λi<k，则正确分类；同样，如果μj<k，则ξj=0和vj正确分类。我们有以下分类：

1. 如果0<λi<k，则ui在边缘，分类正确。同样，如果0<μj<k，则Vj在边缘，并且分类正确。
2. 如果λi=k，那么如果1，点ui可以正确分类，或者它位于正确侧的边缘内，但是如果1，那么它被错误分类。同样，如果μj=k，则如果ξj≤1，则点vj可正确分类，或位于正确侧的边缘内，但如果ξj>1，则其分类错误。
3. 如果λi=0，则ui被正确分类。同样，如果μj=0，则Vj分类正确。

如果原始解w=06，那么方程



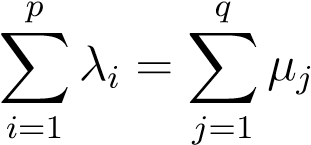
意味着要么存在一些指数i0，使得λi0>0，要么存在一些指数j0，使得μj0>0。约束条件

PQ

十倍

λi=μj i=1 j=1

意味着存在一些指数i0，使得λi0>0，并且存在一些指数j0，使得μj0>0。然而，先验的，没有什么能阻止所有非零的λi的λi=k或所有非零的μj的μj=k的情况。如果发生这种情况，我们可以用更大的k值重新运行优化方法。观察方程



意味着如果存在指数I0，如0<λI0<k，则存在指数J0，如0<μj0<k，反之亦然。如果下面的温和假设成立，那么可以找到b。

（SVMS2）的标准裕度假设。有一些指数i0，如0<λi0<k，有一些指数j0，如0<μj0<k。这意味着一些ui0被正确分类，在蓝边上，一些vj0被正确分类，在红边上。

如果（svms2）的标准裕度假设成立，那么=0和礹j0=0，那么我们就得到了活动方程。

w>ui0−b=1和−w>vj0+b=1，

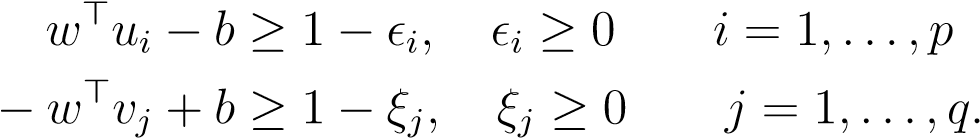
我们得到

.

备注：有一个廉价版本的问题（SVM）s2），它将术语（1/2）w>w从目标函数中删除：

软边缘分类器（SVMS2L）：

最小化

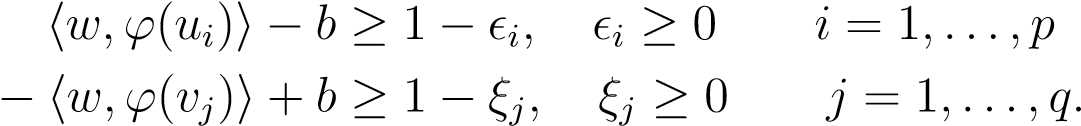


上面的程序是一个线性程序，它可以最小化错误分类的点的数量，但不关心执行最小界限。Boyd和Vandenberghe给出了其使用示例；见[29]第8.6.1节。

问题（SVMS2）的“kernelized”版本如下：

软边缘内核SVM（SVMS2）：

最小化



在重新计算对偶函数时，我们发现对偶程序是由

最小化



其中，k是第54.1节末尾给出的``核对称矩阵（其中`=p+q）。我们也发现

，

所以

，

以及分类功能

f（x）=sgn（hw，（x）i-b）

由给出

.

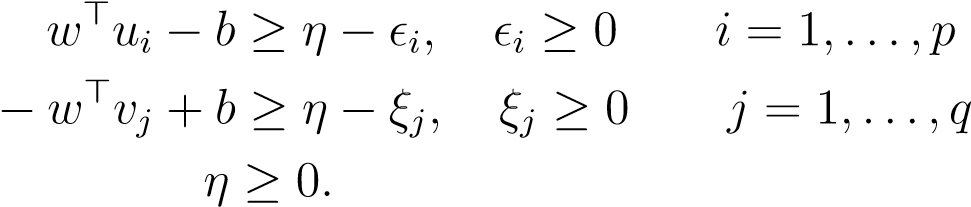
## 54.3软边界支持向量机；（SVMS20）

在这一节中，我们考虑问题（svms2）的泛化，对于来自问题（svmh2）的软边界svm的一个版本，通过增加额外的自由度，即代替边缘δ=1/kwk，我们使用边缘δ=η/kwk，其中η是我们希望最大化的某个正常数。为此，我们在目标函数中添加一个术语−kmη。

（1/2）w>w以及“正则化项”，其目的是使和ξ稀疏，其中km>0和ks>0是固定常数，可以调整以确定η和正则化项的影响。

软利润SVM（SVMS20）：

最小化



SVM问题的这一版本首先在scho–lkopf、smola、williamson和bartlett[143]中以ν-svc（或ν-svm）的名义进行了讨论，也用于sch–olkopf、platt、shawe–taylor和smola[142]。snu-svc方法在scho–lkopf和smola[141]中也有介绍（其中包含更多）。Chan和Lin[36]对第54.2节（有时称为c-svm方法）中所述的方法（即c-svm方法）与ν-svc方法之间的差异进行了彻底的研究。

对于这个问题，不再清楚如果（w，η，b，，ξ）是一个最优解，那么w=06，η>0。事实上，如果点集不是线性可分的，并且Ks选得太大，问题（SVMS20）可能无法得到最优解。

我们表明，为了解决这个问题，我们必须选择km和ks，以便

km≤最小2pks，2qks。

如果我们用

，

则km≤min 2pks，2qks等于

.

引入镟的原因是，镟（p+q）/2可以解释为未能达到边缘η的最大点数。如果集合ui和vj不是线性可分离的，那么我们必须选取ν，以便使方法的ν≥2/（p+q）具有光学解。如果v<3/（p+q），并且至少有三个点被错误分类，那么我们有一些有趣的保证；见54.5号提案和54.6号提案。

我们问题的目标函数是凸的，约束是仿射的。因此，根据定理49.16（2），如果原始问题（SVMS20）有一个最优解，那么对偶问题也有一个解，对偶间隙为零。这并不立即意味着对偶的最优解会产生原始解的最优解，因为定理49.16（1）的假设不成立。

我们证明，如果原问题有一个w=06的最优解（w，η，，ξ，b），那么对偶问题的任何最优解都确定了λ和礹，进而通过方程确定了w。

，（W）

且η≥0。

还有待确定。对偶的解不能直接确定b，η，，ξ，我们不知道确保它们可以被确定的必要和充分条件。我们能做的最好的就是使用KKT条件。

最简单的充分条件是我们称之为

（SVMS20）的标准裕度假设：存在一些I0，例如0<λi0<ks，存在一些μj0，例如0<μj0<ks。这意味着一些ui0被正确地分类，并且在蓝边上，一些vj0被正确地分类，并且在红边上。

在这种情况下，通过互补松弛度，可以证明=0，并且相应的不等式是活动的，也就是我们有方程

w>ui0−b=η，−w>vj0+b=η，

所以我们可以解b和η。然后，由于通过互补松弛度，如果0，则λi=ks，如果ξj>0，则μj=ks，与该0和μj>0对应的所有不等式都是有效的，我们可以解出i和ξj。

如果2/（p+q）≤ν<3/（p+q）且至少三个点被错误分类，那么我们可以保证存在一些i0，使得约束w>ui0−b=η处于活动状态，或者存在一些j0，使得约束−w>vj0+b=η处于活动状态。

如果（w，η，，ξ，b）是w=06的问题（svms20）的最优解，那么点ui和vj分类如下：

=0，则点ui被正确分类，位于蓝色边缘（方程w>x=b+η的超平面hw，b+η）或蓝色边缘的正确侧（蓝色侧）。同样，如果ξj=0，则点vj被正确分类，并位于红边（方程w>x=b−η的超平面hw，b−η）或红边（红边）的正确一侧。

（2）如果为0，则点ui位于边缘（板）内，但位于分离超平面的正确侧（蓝色侧）。如果，那么UI位于分离的超平面上。同样，如果0<ξj≤η，则点vj位于边缘（板）内，但位于分离超平面的正确侧（红色侧）。如果ξj=η，则vj位于分离超平面上。

，则点ui位于分离超平面的错误一侧（红色一侧）；它被错误分类。同样，如果ξj>η，那么点vj位于分离超平面（蓝色面）的错误一侧；它被错误分类。

0）的点称为边缘误差；它们要么位于板内，要么被错误分类。

线性约束由（2（p+q）+1）×（n+p+q+2）矩阵给出，以块形式给出

，

线性约束表示为

.

目标函数由

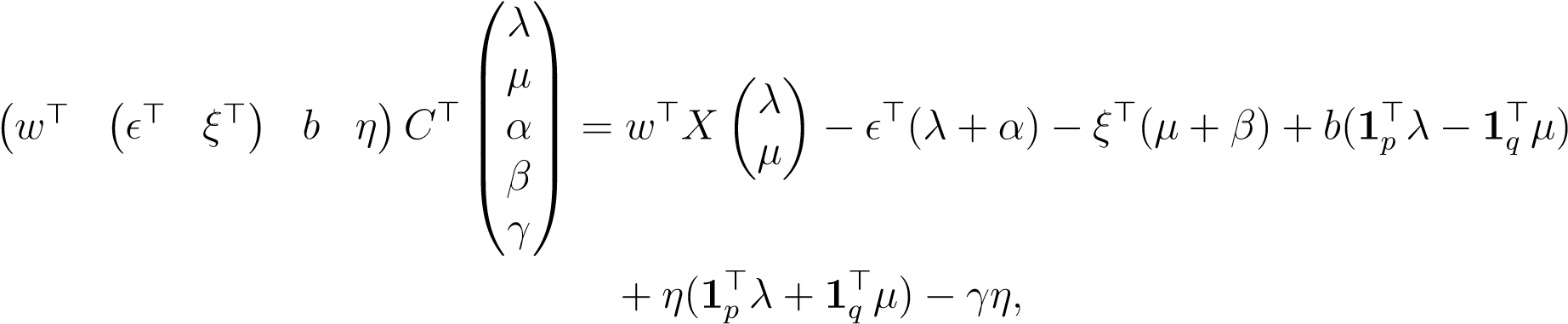
.

给出了，和γ∈r+的拉格朗日

通过

.

自从



拉格朗日可以写成

.

为了找到双函数g（λ，μ，α，β，γ），我们用

关于w，，ξ，b和η。由于拉格朗日是凸的，（r×r，一个凸的开集，根据定理39.11，拉格朗日在（iff=0）中有一个最小值，因此我们计算了它相对于w，，ξ，b，η的梯度，得到

.

通过设置=0，我们得到方程

（W）

λ+α=ks1pμ+β=ks1q

1>pλ=1>qμ，

和

（γ）

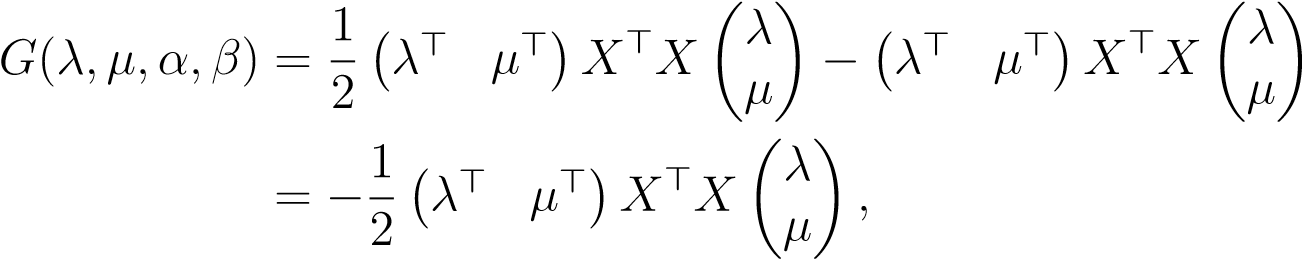
第二和第三个方程等价于箱约束

0≤λi，μj≤ks，i=1，…，p，j=1，…，q，

因为γ≥0方程（γ）等于

.

将w从（w）插回拉格朗日，简化后，我们得到

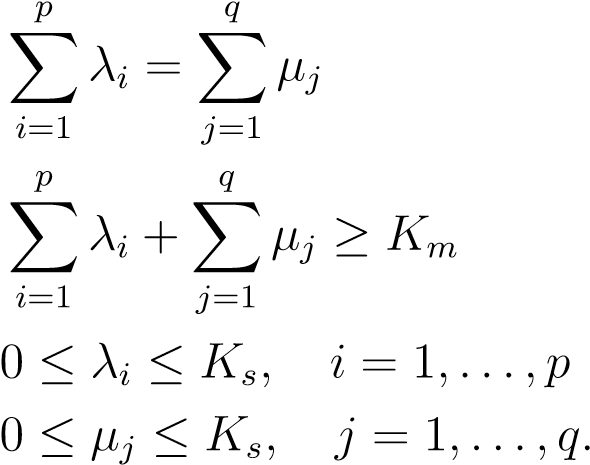


因此，双函数独立于α，β，并由

.

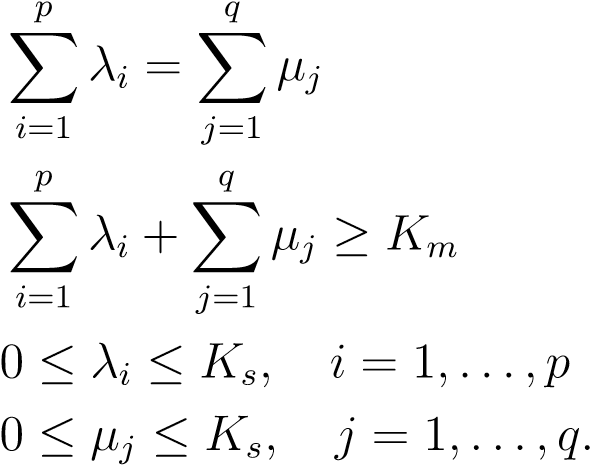
双程序由

最大化受试者



最后，双程序相当于以下最小化程序：

最小化



通过使用基于梯度下降的数值程序（例如第51.6节中的ADMM）来求解双程序。如果原始问题是可解的，则会得到λ和μ的解。一旦得到λ和μ的溶液，我们就得到

.

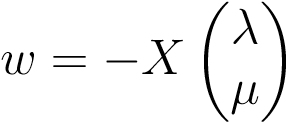
如前所述，定理49.16（2）的假设成立，所以如果原始问题（SVMS20）有一个w=06的最优解，那么对偶问题也有一个解，对偶间隙为零。因此，对于最优解，我们有

，

也就是说

，

从那以后



我们得到

，

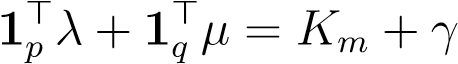
会产生

.（）

因此，η≥0。

评论：

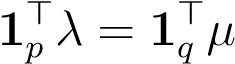
1. 问题的目标函数（SVMS20）是问题的目标函数（SVMS1）的一半，但有些约束是不同的。然而，问题（SVMS20）的主要优点是W总是被确定的。
2. 因为我们证明了如果原始问题（SVMS20）有一个W=06的最优解，那么η≥0，人们可能会奇怪为什么约束η≥0被包括在内。如果我们删除这个约束，很容易看出唯一的区别是，它不是方程



我们得到了方程

.

因为方程式



在第一种情况下，我们获得

（1）

在第二种情况下，我们得到

.（2）

如果η>0，则通过互补松弛度γ=0，在这种情况下（1）和（2）是等效的。但如果η=0，则γ可以严格为正。

不清楚将约束η≥0包含在初始值中的选择是否有利，除非在对偶程序中，方程和不等式可能是有利的。

1>pλ=1>qμ1>pλ+1>qμ≥km

包括而不是方程式

.

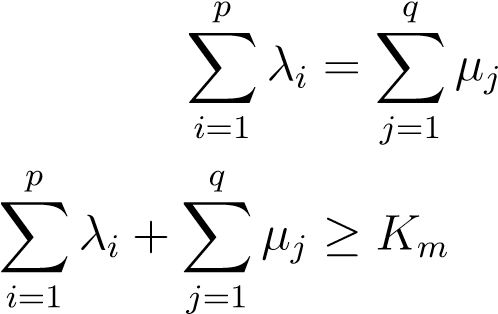
也许使用不平等可以更容易地解决二元问题。为了解决这个问题，我们似乎需要对一些测试数据运行实际的解决方案。

回到问题（SVMS20），互补松弛条件根据λ和μ的值对点进行分类。事实上，我们有

当j=1，…，q时，ξjβj=0。同样，如果λi>0，那么相应的约束是激活的，同样，如果μj>0。既然=0，既然μj+βj=ks，我们有j j i i s，如果ξj>0，那么μj=ks。因此，如果i s=0和ui被正确分类，同样，如果μj<ks，则ξj=0和vj被正确分类。除了约束之外

0≤λi≤ks，0≤μj≤ks，

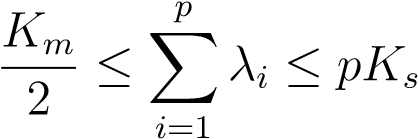
我们也有限制



这意味着

和。（？）

因为λ，μ都是非负的，如果所有i的λi=ks，如果所有j的μj=ks，那么



和

，

因此，除非km≤min 2pks，2qks，这些约束条件是不满足的，因此我们假设km≤min 2pks，2qks。（†）中的方程也意味着存在一些i0，例如λi0>0，一些j0，例如μj0>0。

我们有以下分类（记住，η>0）：

1. 如果0<λi<ks，则ui在边缘，分类正确。同样，如果0<μj<ks，则Vj在边缘，并且分类正确。
2. 如果λi=ks，那么我们不看=0就不能说更多，那么点ui在边界上，并且分类正确；如果0，那么ui在正确的边上的边界内，但是如果它被错误分类。同样，如果μj=ks，那么我们不看ξj就不能说更多。如果ξj=0，那么点vj在边缘上，分类正确，如果0<ξj≤η，那么vj在正确的边上的边缘内，但是如果ξj>η，那么它被错误分类。
3. 如果λi=0，则ui被正确分类。同样，如果μj=0，则Vj分类正确。无法判断UI是否在页边空白处，同样对于VJ也是如此。

我们发现定义v>0很方便，这样

km=（p+q）ksν，

那就是

，

因此，目标函数）由

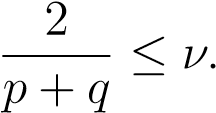
，

k=（p+q）ks，所以km=kν，ks=k/（p+q）。

观察条件km≤min 2pks，2qks等于

，

条件ks≤km/2等于



由于我们通过用一个公共的正因子重新标定来获得一个等价的问题，因此可以方便地将ks归一化为

，

在这种情况下，km=ν。这种方法被称为nv-支持向量机。

在（svms20）的标准裕度假设下，存在一些i0，例如0<λi0<ks，一些j0，例如0<μj0<ks，并且由互补松弛条件=0和ξj0=0，因此我们有两个主动约束。

w>ui0−b=η，−w>vj0+b=η，

我们可以解b和η，得到

.

方程（†）和盒子不等式

0≤λi≤ks，0≤μj≤ks

也意味着以下事实：

提案54.1.如果问题（SVMS20）的最优解w=06且η>0，则以下事实成立：

1. 至多为ν（p+q）/2点ui未达到边缘η，至多为ν（p+q）/2点vj未达到边缘η。
2. 其中，至少ν（p+q）/2点ui的裕度最大为η，至少ν（q+q）/2点的裕度最大为η。

证据。（1）回想一下，对于w=06且η>0的最优解，我们得到γ=0，因此通过（γ），我们得到了方程。

而且。

如果ui未能达到边缘η，则为0，并通过补充松弛度λi=ks=km/（ν（p+q）），因此，如果存在此类点，则

，

所以

.

如果VJ未能达到裕度η，用pqj=1μj代替（其中qf是未能达到裕度η的VJ点数），则采用类似的推理。

（2）点ui的最大裕度为ηiffλi>0。如果

im=i∈1，…，pλi>0和pm=im，

然后

，

由于λi≤ks=km/（ν（p+q）），我们得到

，

会产生

.

如果一个点vj的最大裕度为η，则采用类似的推理。

请注意，如果选择了ν，使之ν<2/（p+q），那么ν（p+q）/2<1，这意味着没有一个数据点被错误分类；换句话说，uis和vjs是线性可分的。因此，我们再次看到，如果uis和vjs不是线性可分离的，我们必须选择v，使2/（p+q）≤v≤min 2p/（p+q），2q/（p+q）方法成功。

下面的命题阐明了常数v在建立保证金宽度和保证金误差点数量之间的权衡中的作用。特别地，它表明如果问题（SVMS20）有一个w=06的最优解，并且如果v<min 2p/（p+q），2q/（p+q），那么至少一些UI或一些VJ被正确分类。显然我们有2/（p+q）≤min 2p/（p+q），2q/（p+q）。

提案54.2.假设是问题（SVMS20）的一个最优解，其中w=06且η>0，设pf为误分类的点Ui个数（qf为误分类的点Vj个数（ξj>0）。如果pf+qf≥3且如果2/（p+q）≤ν<（pf+qf）/（p+q），那么要么存在一些i，使得约束w>ui−b=η处于活动状态，要么存在一些j，使得ξj=0且约束−w>vj+b=η处于活动状态。

证据。（1）我们可以假设ks=1/（p+q）。我们自相矛盾。因此，我们假设，对于all=0，那么约束w>ui−b≥η是不活动的，即w>ui−b>η，对于所有j∈1，…，q，如果ξj=0，那么约束−w>vj+b≥η是不活动的，即−w>vj+b>η。

设，并设pf=i和qf=j（当然，η>0）。

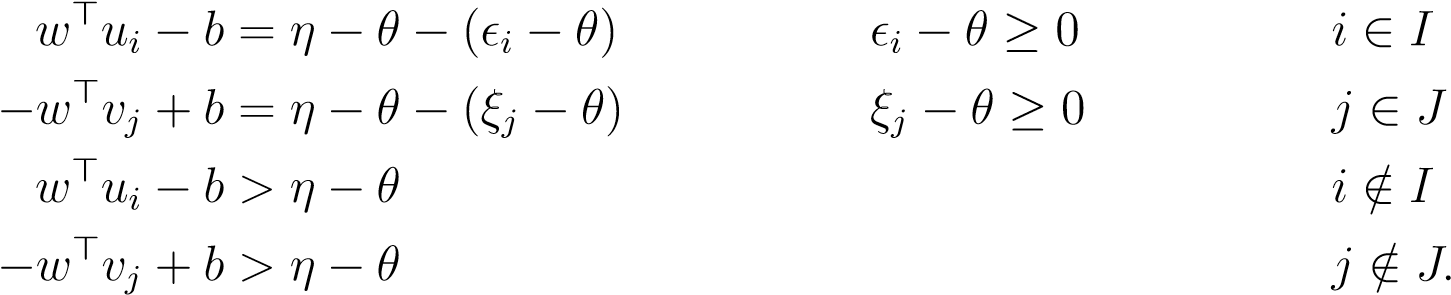
假设pf+qf≥3。由于互补松弛性，所有约束i∈i和j∈j都是主动的，因此我们的假设是



对于任何θ>0，这样

，

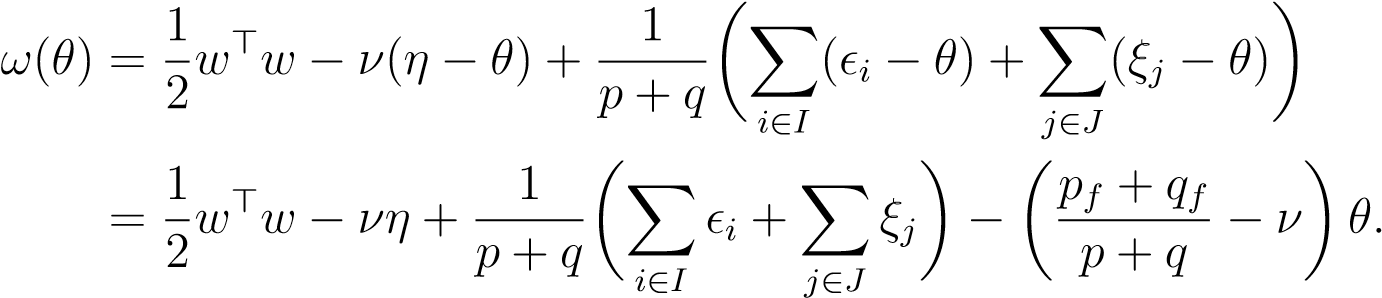
我们可以写信



目标函数的初始值是

，

新的价值是



因为根据假设pf+qf≥3，如果

，

那么涉及θ的项是负的，所以

ω（θ）<ω（0），

通过选择θ，我们得到了一个可行的解，与解的最优性（对于向量）相矛盾。

，与ξ−θ类似。

请注意，如果pf+qf=p+q且ν<min 2p/（p+q），2q/（p+q）≤1，则命题54.5产生矛盾。因此，pf+qf<p+q，即至少一些ui或一些vj是

分类正确

注：如果集合ui和vj是线性可分的，那么从定理49.12我们可以知道，一些ui在蓝边上，一些vj在红边上。

我们也有下面的命题，它给出了一个充分的条件，暗示η和b可以从对偶的最优解（λ，μ）中找到。

提案54.3.是w=06且η>0的问题（SVMS20）的最优解，如果2/（p+q）≤ν<4/（p+q）和pf，qf≥2，则始终可以从对偶的最优解（λ，μ）中确定η和b。

证据。由于pf+qf≥4，根据命题54.5，要么有一些i0，例如=0，约束w>ui0−b=η处于活动状态，要么有一些j0，例如ξj0=0，约束−w>vj0+b=η处于活动状态，正如我们已经解释的那样，问题（svms20）满足具有零dua的条件。间隙。因此，对于最优解，我们有

，

也就是说

，

从那以后

，

我们得到

.（）

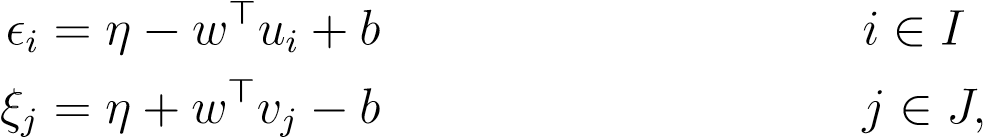
设j=j∈1，…，qξj>0。假设i≥2和j≥2。我们知道，所有i∈i的λi=1/（p+q），所有j∈j的λj=1/（p+q），因此以下方程是有效的：

i∈i j∈j。

但是（）可以写为

，（）

从那以后



通过代入方程（），我们得到

.

我们还知道w>ui0−b=η或−w>vj0+b=η。在第一种情况下，b=−η+w>ui0，通过在上述方程中代入b，我们得到一个形式的方程。

，

也就是说，

在第二种情况下b=η+w>vj0，我们得到一个形式的方程

，

也就是说，

我们需要选择，使2 i/（p+q）−ν=06和2 j/（p+q）−ν=06。由于i≥2和j≥2，因此，如果v<4/（p+q），则会出现这种情况。如果满足这个条件，我们可以求出η，然后我们从b=−η+w>ui0或b=η+w>vj0中找到b。

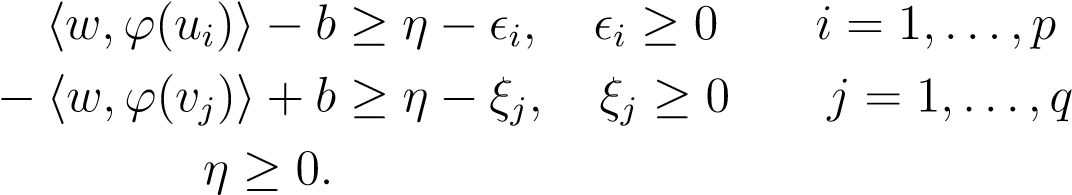
注：如果集合ui和vj是线性可分的，那么从定理49.12我们可以知道，一些ui在蓝边上，一些vj在红边上，因此可以确定b和δ。虽然我们可以确保某些UI被正确分类或某些VJ被正确分类，但似乎不可能证明相应的约束在没有附加假设（如pf+qf≥3）的情况下是活动的。

支持向量机的优点之一是有助于在vc维（vapnik和chernovenkis发明的概念）方面找到有趣的统计界限。我们不会在这里讨论这个问题，而是将读者引向Vapnik[176]（尤其是第4章和第9-13章）。

问题（SVMS20）的“kernelized”版本如下：

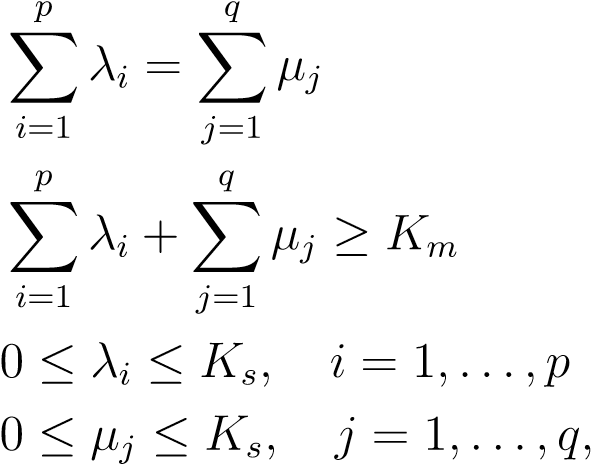
软边缘内核SVM（SVMS20）：

最小化



通过对双程序的推导，我们得到

最小化



其中k是第54.1节的核矩阵。

如第54.2节所述，我们获得

，

所以

，

以及分类功能

f（x）=sgn（hw，（x）i-b）

由给出

.

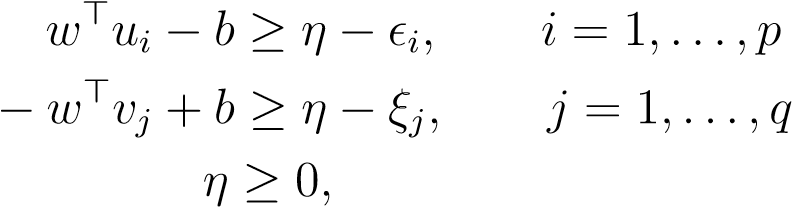
## 54.4软保证金SVM；（SVMS3）

在这一部分中，我们考虑问题（SVMS20）的版本，在该版本中，我们不使用函数作为正则化函数，而是使用二次函数



软利润SVM（SVMS3）：

最小化

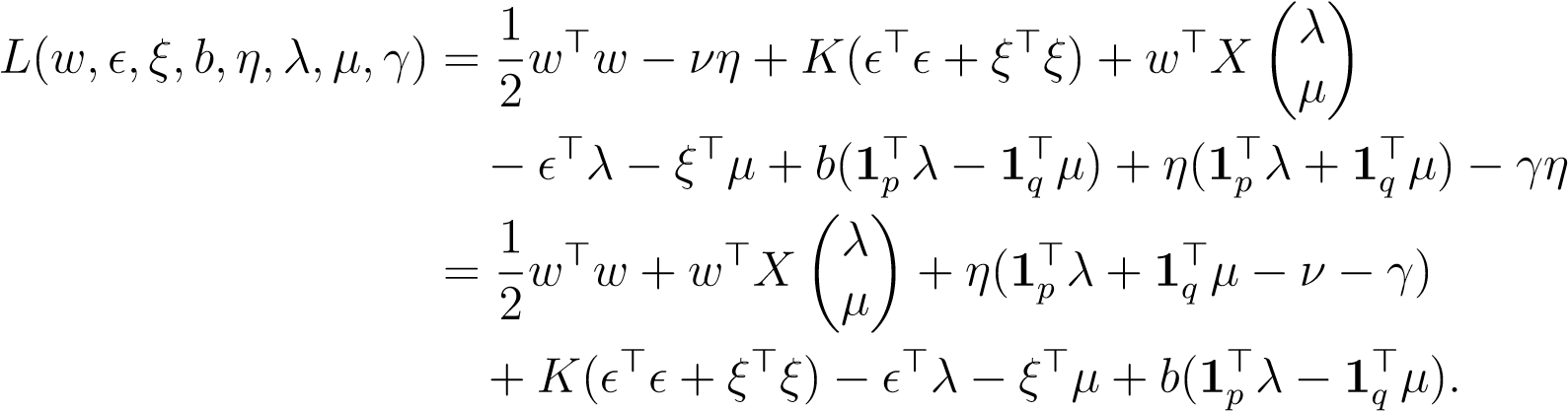


式中，ν和k是两个给定的正常数。如前所述，选择k=1/（p+q）比较方便。

这个问题公式的新扭曲是，如果为0，则相应的不等式意味着通过将i设置为零而减少的值得到的不等式w>ui−b≥η，同样，如果ξj<0，则相应的不等式−w>vj+b≥η−ξj意味着in等式−w>vj+b≥η，通过将ξj设置为零，同时减小kξk2的值得到。因此，如果（w，b，，ξ）是问题（svms3）的最优解，则无需将松弛变量限制为非负变量，这就简化了一些问题。

这种方法的一个优点是由λ确定，ξ由μ确定。我们也可以省略约束η≥0，因为对于最优解，可以用对偶性表示η≥0。

拉格朗日由

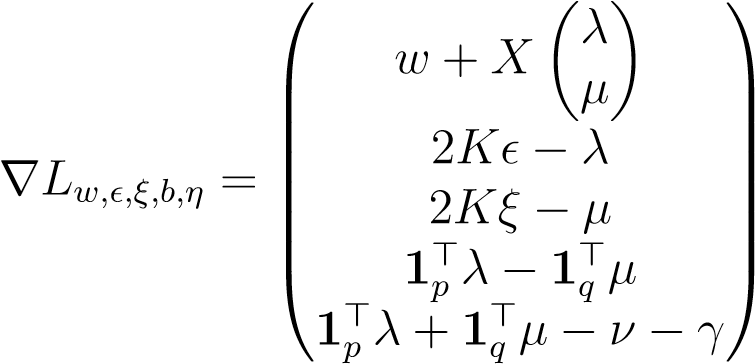


为了找到关于w、ξ、b和η的双函数g（λ、μ、γ），我们将其最小化。因为拉格朗日是凸的（

开集，根据定理39.11，拉格朗日在（=0，

54.4。软保证金SVM；（SVMS3）

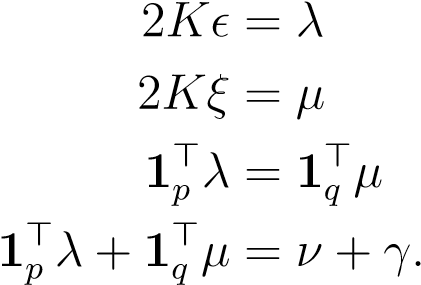
所以我们计算。梯度由下式给出



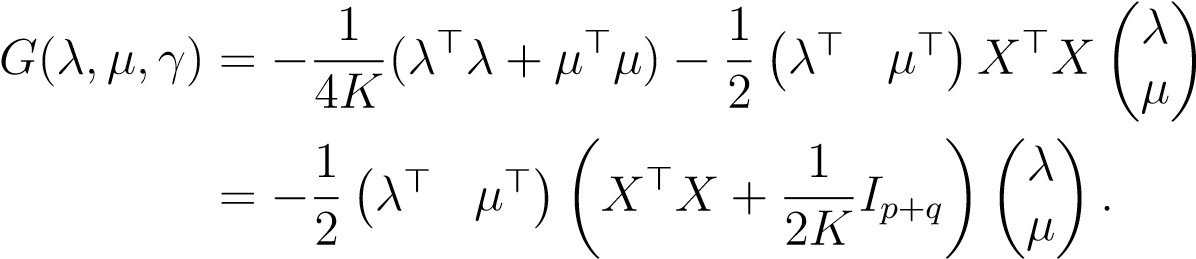
通过设置=0，我们得到方程

（W）

和

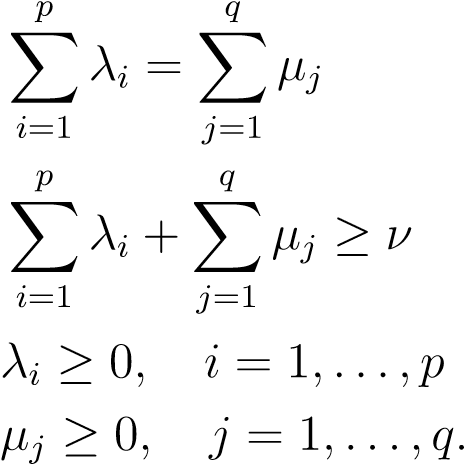


最后两个方程与问题（SVMS20）中得到的最后两个方程相同。我们可以用其他方程得到双函数g（λ，μ，γ）的下列表达式。



因此，双程序相当于最小化程序。

最小化



上述程序与为问题（SVMS20）而得到的程序相似，但矩阵x>x被矩阵x>x+（1/2k）ip+q所代替，该矩阵自k>0起为正定，且不再存在λi≤k和μj≤k的不等式。然而，这些限制意味着存在一些i0，例如λi0>0，而一些j0，例如μj0>0。

通过使用基于梯度下降的数值程序（例如第51.6节中的ADMM）来求解双程序。如果原始问题是可解的，则会得到λ和μ的解。我们从问题（SVMS20）中的λ和μ以及γ中获得w；即，

p q w=xλiui−xμjvj。

I=1 J=1

由于变量i和μj不被限制为非负，我们不再有涉及它们的互补松弛条件，但我们知道

.

也因为约束

和

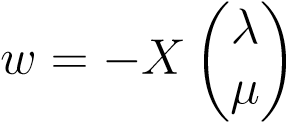
意味着存在一些i0，例如λi0>0和一些j0，例如，μj0>0，我们有且ξj0>0，这意味着至少两个点被错误分类，因此问题（svms3）只应在集ui和vj不可线性分离时使用。我们可以使用与任何i0对应的主动约束来解b和η，从而使λi0>0和任何j0，从而使μj0>0，我们得到

.

我们也可以利用最优性间隙为0的事实求出η。我们有

，

从那以后



我们得到

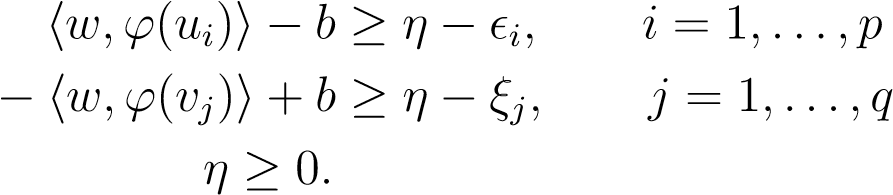
.

上述结果证实，在最优性条件下，我们得到η≥0。

问题（SVMS3）的“kernelized”版本如下：

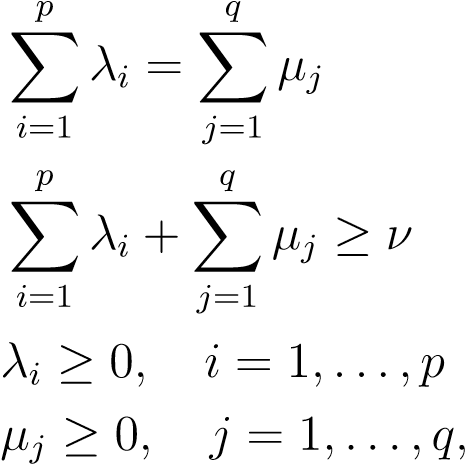
软边缘内核SVM（SVMS3）：

最小化



通过对对偶程序的推导，我们得到

最小化



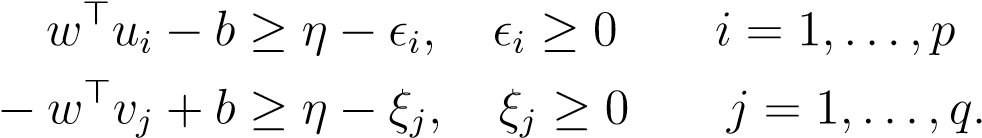
其中k是第54.1节的核矩阵。然后，w、b和f（x）的计算与第54.3节中的计算完全相同。

## 54.5软边界支持向量机；（SVMS4）

在本节中，我们通过在目标函数中添加术语（1/2）b2来考虑问题（SVMS20）的变化。结果表明，在将拉格朗日最小化的情况下，不仅可以求出双函数G，而且还可以求出B。我们还抑制了约束η≥0，这是多余的。

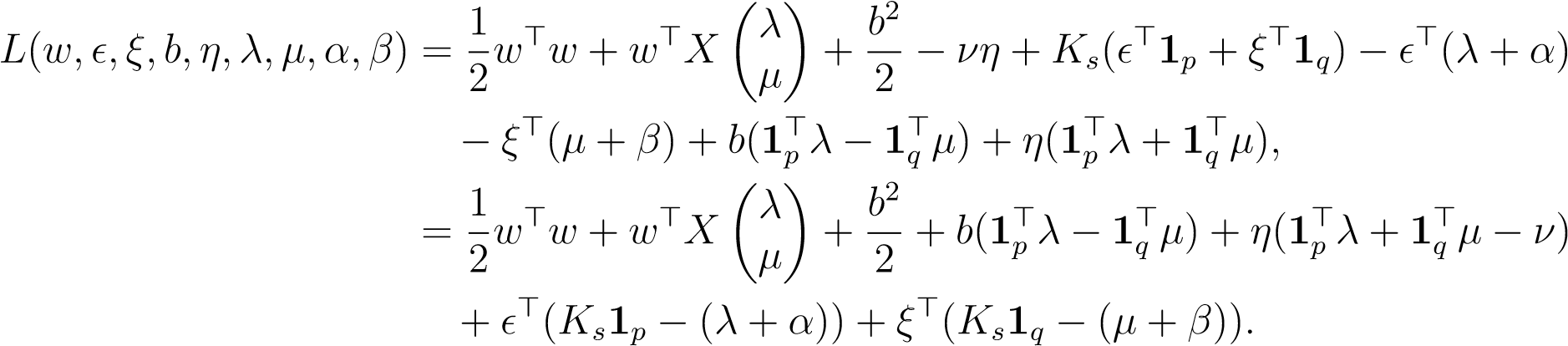
软利润SVM（SVMS4）：

最小化



为了简化演示，我们假设k=1，并将k写为1/（p+q）。

拉格朗日）的



为了找到关于的双函数g（λ，μ，α，β），我们将其最小化。由于拉格朗日是凸的（，一个凸的开集），根据定理39.11，拉格朗日在（w，，ξ，b，η）iff=0中有一个最小值，因此我们计算了它相对于w，，ξ，b，η的梯度，得到

.

通过设置=0，我们得到方程

（W）

和

.（b）

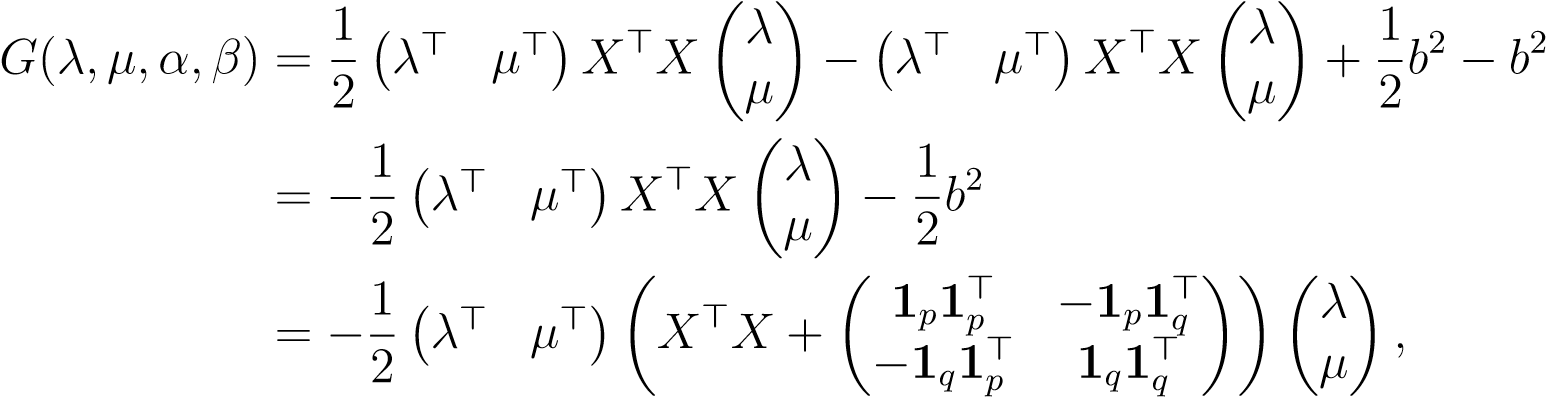
第二和第三个方程等价于箱约束

0≤λi，μj≤ks，i=1，…，p，j=1，…，q.

因为我们假设原始问题有一个w=06的最优解，我们有

.

把w从（w）和b从（b）插回拉格朗日，我们得到



因此，双函数独立于α，β，并由

.

双程序由

最大化受试者

PQ

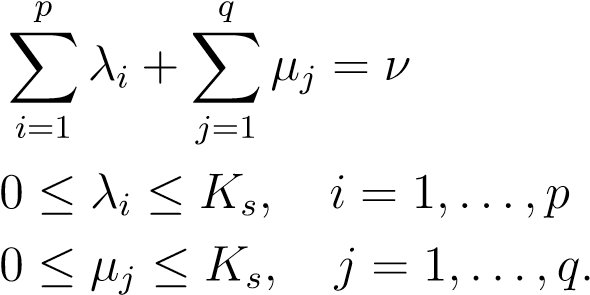
十倍

λi+μj=νi=1 j=1 0≤λi≤ks，i=1，…，p

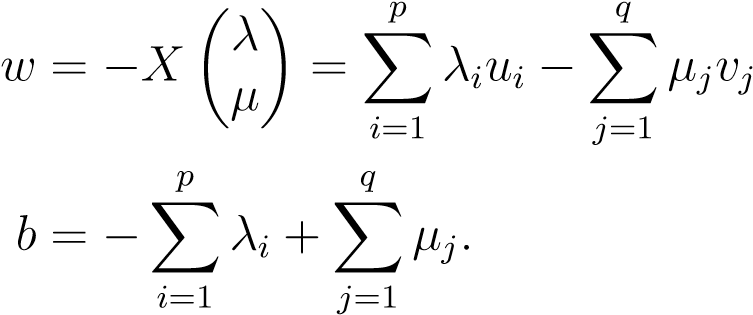
0≤μj≤ks，j=1，…，q.

最后，双程序相当于以下最小化程序：

最小化



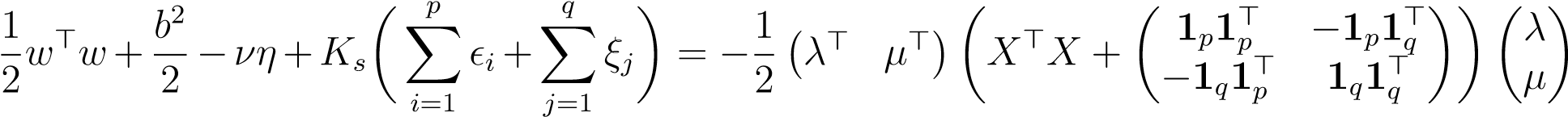
通过使用基于梯度下降的数值程序（例如第51.6节中的ADMM）来求解双程序。如果原始问题是可解的，则会得到λ和μ的解。一旦得到λ和μ的溶液，我们就得到



如前所述，定理49.16（2）的假设成立，所以如果原始问题（SVMS4）有一个w=06的最优解，那么对偶问题也有一个解，对偶间隙为零。因此，对于最优解，我们有

，

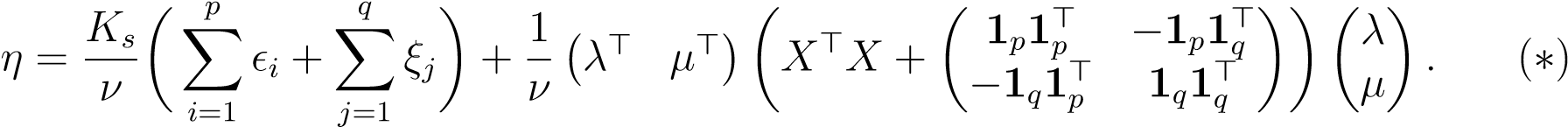
也就是说



从那以后

，

我们得到



自从

-

是正半定的，因此我们确定η≥0。

因为ks=1/（p+q），为了满足约束条件



并且0≤λi，μj≤1/（p+q）要满足要求，我们必须

ν≤1.

方程式

PQ

十倍

λi+μj=νi=1 j=1

也意味着要么存在一些i0，使得λi0>0，要么存在一些j0，使得μj0>0。

在（svms4）的标准裕度假设下，要么存在一些i0，例如0<λi0<ks，要么存在一些j0，例如0<μj0<ks，并且通过互补松弛条件=0，因此我们得出

w>ui0−b=η，或−w>vj0+b=η，

我们可以求出η。

方程（†）和盒子不等式

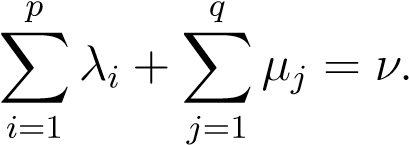
0≤λi≤ks，0≤μj≤ks

也意味着以下事实：

提案54.4.如果问题（SVMS4）的最优解w=06且η>0，则以下事实成立：

1. 在最大的ν（p+q）点，ui和vj未能达到边缘η。
2. Ui和Vj上至少有ν（p+q）点的边缘至多为η。

证据。（1）回想一下，对于w=06且η>0的最优解，我们得到了方程



如果ui未能达到裕度η，则为0，并通过补充松弛度λi=ks=1/（p+q）。同样，如果VJ未能达到边缘，则ξj>0，并通过补充松弛度μj=ks=1/（p+q）。假设pf点ui不符合边际，qf点vj不符合边际。然后

，

所以pf+qf≤ν（p+q）。

（2）一点ui的裕度最大为ηiffλi>0，一点vj的裕度最大为ηiffμj>0。如果

im=i∈1，…，pλi>0和pm=im|

和

jm=j∈1，…，qμj>0和qm=jm|

然后

，

由于λi，μj≤ks=1/（p+q），我们得到

，

得出pm+qm≥ν（p+q）。

请注意，如果选择了ν，则使之等于ν<1/（p+q），则表示没有一个数据点被错误分类；换句话说，uis和vjs是线性可分离的。因此，我们发现，如果uis和vjs不是线性可分离的，我们必须选择v，这样1/（p+q）≤v≤1，方法才能成功。

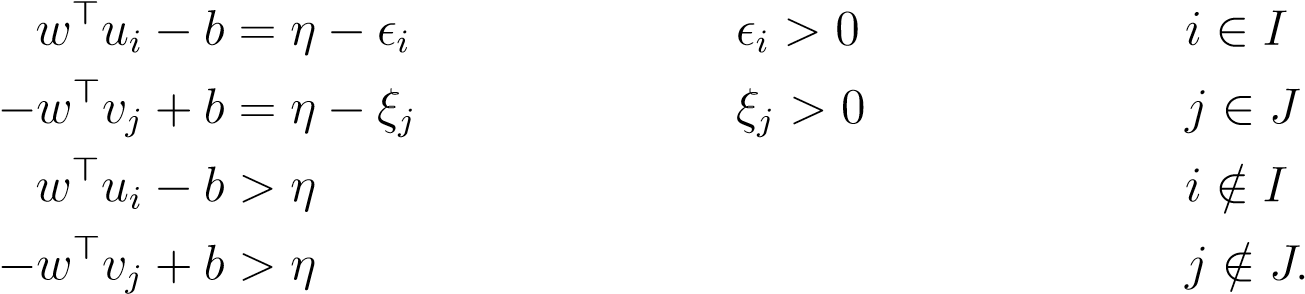
下面的命题阐明了常数v在建立保证金宽度和保证金误差点数量之间的权衡中的作用。特别地，它表明，如果问题（SVMS4）有一个w=06且η>0的最优解，并且如果ν<1，那么至少一些ui或一些vj被正确分类。显然我们有1/（p+q）≤1。

提案54.5。假设是问题（SVMS4）的最优解，其中w=06，η>0，设pf为误分类的点Ui个数（qf为误分类的点Vj个数（ξj>0）。如果pf+qf≥2且如果1/（p+q）≤ν<（pf+qf）/（p+q），那么要么存在一些i，使得约束w>ui−b=η处于活动状态，要么存在一些j，使得ξj=0且约束−w>vj+b=η处于活动状态。

证据。（1）我们可以假设ks=1/（p+q）。我们自相矛盾。因此，我们假设，对于all=0，那么约束w>ui−b≥η是不活动的，即w>ui−b>η，对于所有j∈1，…，q，如果ξj=0，那么约束−w>vj+b≥η是不活动的，即−w>vj+b>η。

设，并设pf=i和qf=j（当然，η>0）。

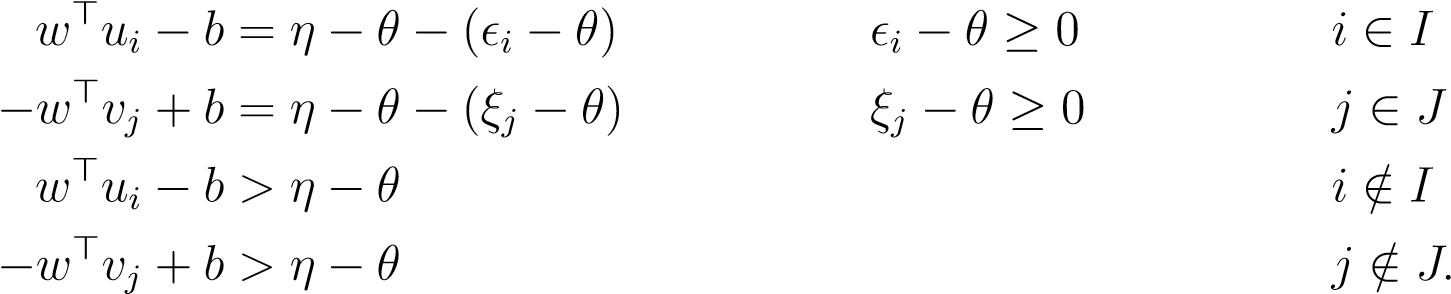
假设pf+qf≥2。由于互补松弛性，所有约束i∈i和j∈j都是主动的，因此我们的假设是



对于任何θ>0，这样

，

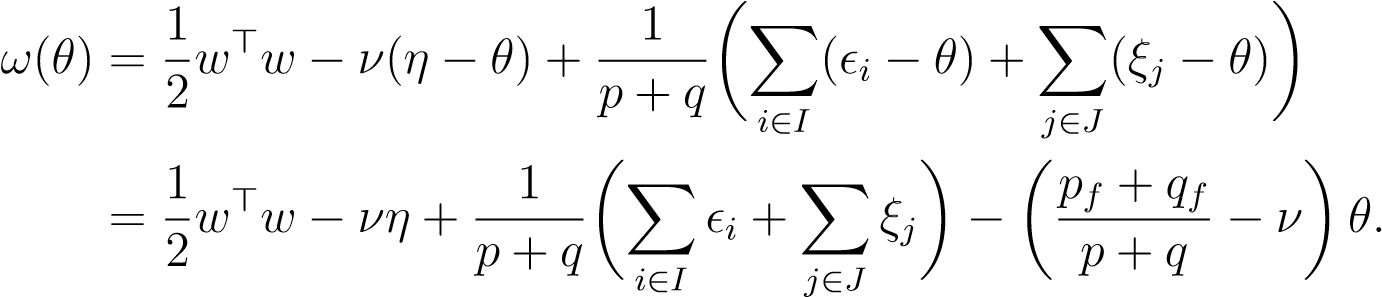
我们可以写信



目标函数的初始值是

，

新的价值是



因为根据假设pf+qf≥2，如果

，

那么涉及θ的项是负的，所以

ω（θ）<ω（0），

通过选择θ，我们得到了一个可行的解，与解的最优性（对于向量）相矛盾。

，与ξ−θ类似。

请注意，如果pf+qf=p+q且ν<1，则54.5命题产生矛盾。因此，pf+qf<p+q，即至少某些ui或某些vj被正确分类。

注：如果集合ui和vj是线性可分的，那么从定理49.12我们可以知道，一些ui在蓝边上，一些vj在红边上。

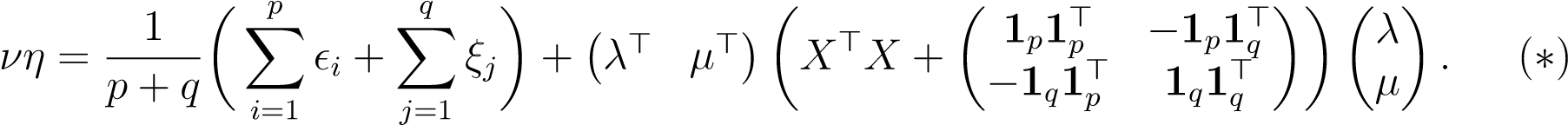
我们也有下面的命题，它给出了一个充分的条件，即η可以从对偶的最优解（λ，μ）中找到。

提案54.6.如果是w=06且η>0的问题（SVMS4）的最优解，如果1/（p+q）≤ν<2/（p+q）且pf+qf≥2，则始终可以从对偶的最优解（λ，μ）中确定η。

证据。正如我们已经解释过的，问题（SVMS4）满足具有零对偶间隙的条件。因此，对于最优解，我们有

，

也就是说



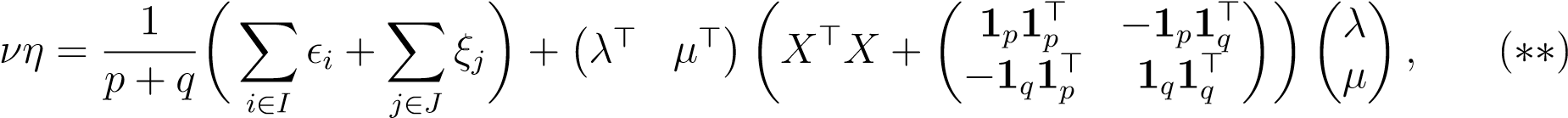
那么，让我们

.

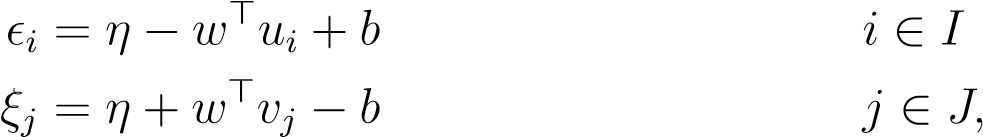
假设i+j≥2。那么我们知道，所有i∈i的λi=1/（p+q），所有j∈j的λj=1/（p+q），因此以下方程是有效的：

i∈i j∈j。

但是（）可以写为



从那以后



通过代入方程（），我们得到

.

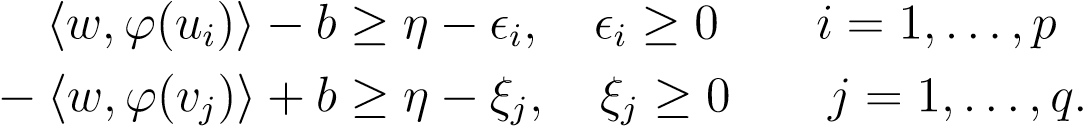
我们可以解出我们需要选择+j≥2，这将是这样的情况，如果\_使（i 1+/（pj+）/q（）p≤+ν<q）−2/（\_p=06+q）。如果这个条件满足，我们假设η。

注：49.12如果setsui在蓝边上，有些ui和vj是线性可分的，那么我们从定理mvj知道它在红边上，所以b和δ可以确定。虽然我们可以确保某些UI被正确分类或某些VJ被正确分类，但似乎不可能证明相应的约束在没有附加假设的情况下是活动的（例如pf+qf≥2）。

问题（SVMS4）的“kernelized”版本如下：

软边缘内核SVM（SVMS4）：

最小化



54.6。软保证金SVM；（SVMS5）

通过对双程序的推导，我们得到

最小化

PQ

十倍

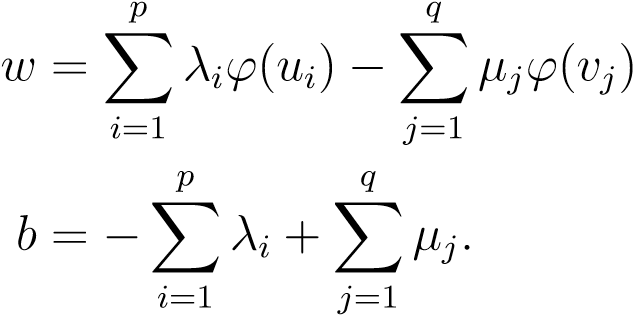
λi+μj=νi=1 j=1

0≤λi≤ks，i=1，…，p

0≤μj≤ks，j=1，…，q，

其中k是第54.1节的核矩阵。

我们得到



分类功能

f（x）=sgn（hw，（x）i-b）

由给出

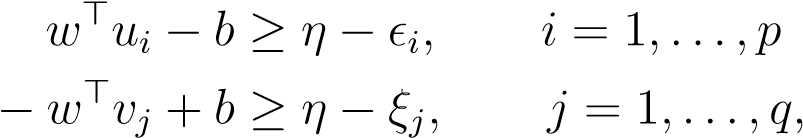
.

## 54.6软保证金SVM；（SVMS5）

在本节中，我们考虑问题（SVMS3）的版本，在该版本中，我们将术语（1/2）b2添加到目标函数中。我们还降低了约束η≥0，这是多余的。

软利润SVM（SVMS5）：

最小化



式中，ν和k是两个给定的正常数。如前所述，选择k=1/（p+q）比较方便。

拉格朗日由

.

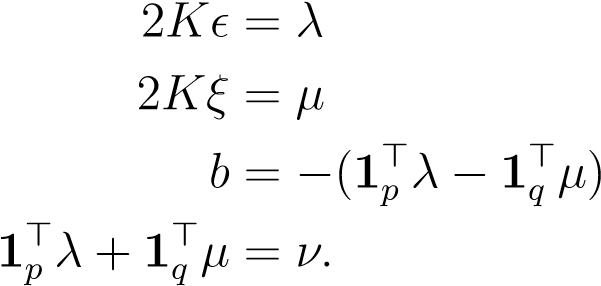
为了找到关于和η的双函数g（λ，μ），我们将其最小化。由于拉格朗日是凸的（，一个凸的开集，根据定理39.11，拉格朗日在（=0）中有一个最小值，所以我们计算。梯度由下式给出



通过设置=0，我们得到方程

（W）

和

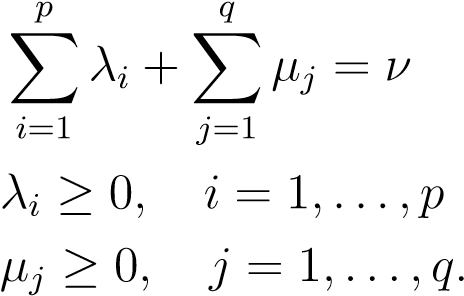


最后两个方程与问题（SVMS4）中得到的最后两个方程相同。我们可以用其他方程得到双函数g（λ，μ，γ）的下列表达式。

.

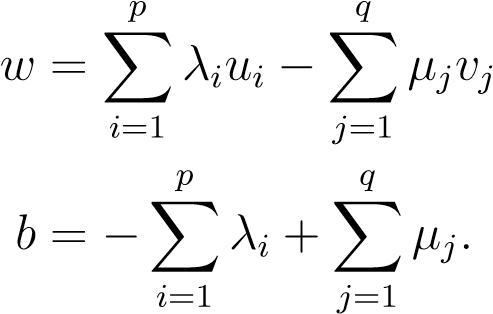
54.6。软保证金SVM；（SVMS5）

因此，双程序相当于最小化程序



通过使用基于梯度下降的数值程序（例如第51.6节中的ADMM）来求解双程序。如果原始问题是可解的，则会得到λ和μ的解。

约束意味着要么存在一些i0，使得λi0>0，要么存在一些j0，使得μj0>0。我们从λ和μ中得到w和b，如问题（svms4）；即，



由于变量i和μj不被限制为非负，我们不再有涉及它们的互补松弛条件，但我们知道

.

也因为约束

PQ

十倍

λi+μj=νi=1 j=1

意味着要么存在一些i0，比如λi0>0，要么存在一些j0，比如μj0>0，我们有0，这意味着至少有一个点被错误分类，所以只有当集合ui和vj不可线性分离时，才应使用问题（svms5）。我们可以使用与任何i0对应的主动约束来求解η，这样λi0>0或任何j0，这样μj0>0。

我们也可以利用最优性间隙为0的事实求出η。我们有

，所以我们得到

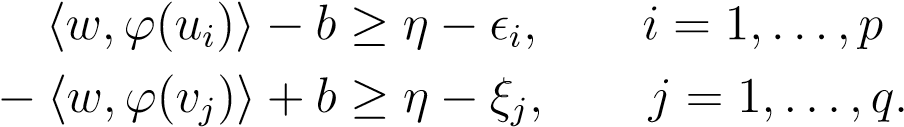
.

上述结果证实，在最优性条件下，我们得到η≥0。

问题（SVMS5）的“kernelized”版本如下：

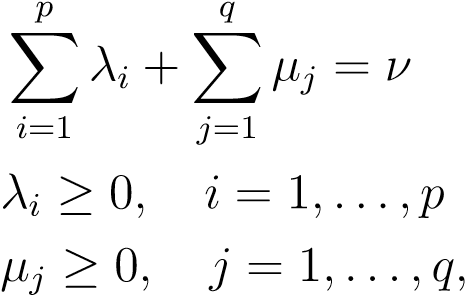
软边缘内核SVM（SVMS5）：

最小化



通过对双程序的推导，我们得到

最小化



其中k是第54.1节的核矩阵。然后，w、b和f（x）的计算与第54.5节中的计算完全相同。

## 54.7支持向量机方法总结与比较

在本章中，我们考虑了六个变量来解决两组点的软边界二值分类问题，并使用支持向量分类方法。目的是找到方程w>x-b=0的分离超平面hw，b。我们还试图找到两个“边缘超平面”hw，方程w>x−b−δ=0和hw，方程w>x−b+δ=0的b−δ，这样δ尽可能大，但错误分类点的数量是最小化的，这是通过允许每个点ui的误差0来实现的，在这个意义上，tHE约束



在约束条件下，每个点vj的误差ξj≥0

−w>vj+b≥δ−ξj

应该保持。

目的是设计一个目标函数，使δ最小化和ξ最大化。对于W和B，优化问题也应该解决，因此必须在W上施加一些约束。另一个目标是尝试使用双程序来解决优化问题，因为解决方案涉及内部产品，因此该问题可以用kern进行泛化。EL功能。

第一次尝试，即使用目标函数



而约束w>w≤1并不能很好地发挥作用，因为这个约束需要用拉格朗日乘子γ≥0来保护，因此，将拉格朗日L最小化以找到对偶函数g，给出了求解形式w的方程。

，

但如果集合不可线性分离，则γ=0可能出现最优解，在这种情况下，无法确定w。这是第54.1节中考虑的问题（SVMS1）。

软利润SVM（SVMS1）：

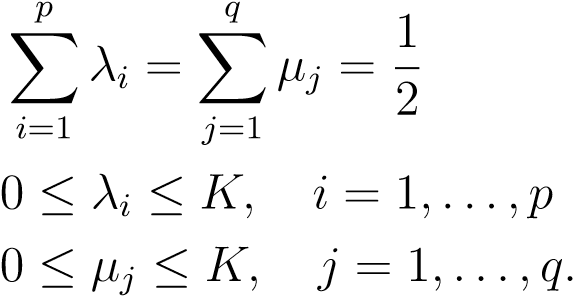
最小化



习惯上写`=p+q。

如第54.1节所示，双程序相当于以下最小化程序：

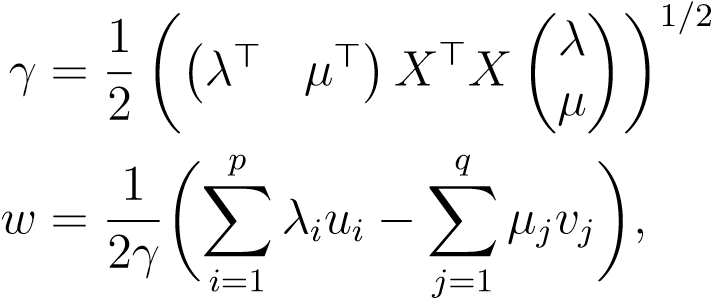
最小化



注意，约束意味着必须选择k，以便

.

如果最佳值为0，则γ=0和=0，因此在这种情况下不可能确定w。但是，如果最佳值大于0，则一旦获得λ和μ的解，我们就可以



所以我们得到

，

如果下面的温和假设成立，则可以找到b和δ。

（SVMS1）的标准裕度假设。有一些指数i0，如0<λi0<k，有一些指数j0，如0<μj0<k。这意味着一些ui0被正确分类，在蓝边上，一些vj0被正确分类，在红边上。

如果（svms1）的标准裕度假设成立，那么=0和祆j0=0，那么我们就有了活动方程。

w>ui0−b=δ和−w>vj0+b=δ，

我们得到b和δ的值

.

第二个更成功的方法是将术语（1/2）w>w添加到目标函数中，并删除约束w>w≤1。然后，根据涉及（线性或二次）的正则化项的选择、边界是如何用delt（隐式地用项1或显式地用项η）进行delt的、以及项（1/2）b2是否添加到目标函数中，该方法有几个变体。

这些方法都具有这样的性质：如果原问题有一个w=06的最优解，那么对偶问题总是确定w，然后在我们称之为标准裕度假设的温和条件下，可以确定b和η。然后可以使用活动的约束来确定。当（1/2）b2添加到目标函数中时，b由方程式确定

.

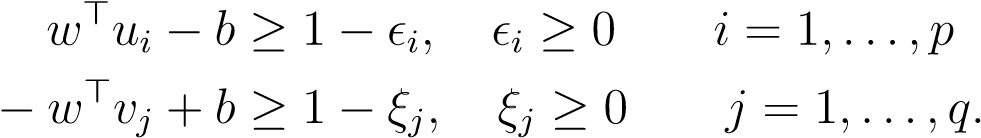
所有这些问题都是凸的，且约束是合格的，所以对偶间隙为零，如果初等有一个w=06的最优解，则得出η≥0。

我们现在更详细地考虑五种变体。

（1）基本软利润支持向量机：（支持向量机2）。

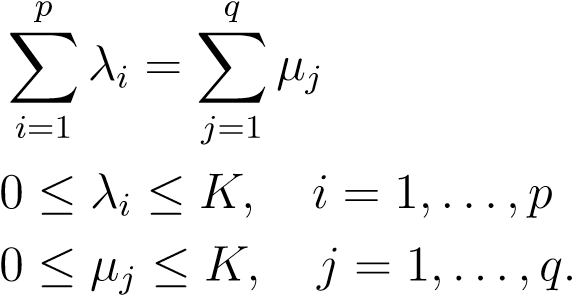
这是一个优化问题，其中正则化项是线性的，裕度δ由δ=1/kwk给出：

最小化



这个问题是所有机器学习或模式分析书籍中讨论的经典问题，例如Vapnik[176]、Bishop[23]和Shawe–Taylor和Christianini[154]。如第54.2节所示，双程序

最小化



我们可以用对偶程序来求解初等问题。一旦发现λ≥0，μ≥0，w由下式得出：

，

但是b不是由二元决定的。

互补松弛条件意味着，如果0，则λi=k，如果ξj>0，则μj=k。因此，如果λi<k，则=0和ui被正确分类，同样，如果μj<k，则ξj=0和vj被正确分类。

先验没有什么能阻止所有非零λi的λi=k或所有非零μj的μj=k的情况。如果发生这种情况，我们可以用更大的k值重新运行优化方法。如果下面的温和假设成立，则可以找到b。

（SVMS2）的标准裕度假设。有一些指数i0，如0<λi0<k，有一些指数j0，如0<μj0<k。这意味着一些ui0被正确分类，在蓝边上，一些vj0被正确分类，在红边上。

如果（svms2）的标准裕度假设成立，那么=0和礹j0=0，那么我们就得到了活动方程。

w>ui0−b=1和−w>vj0+b=1，

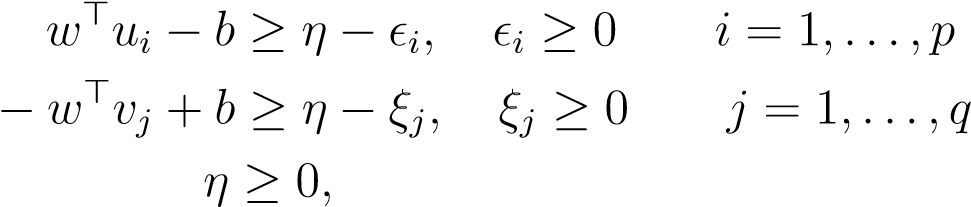
我们得到

.

（2）基本软边挓-SVM问题（SVMS20）。

这个问题的推广（svms2）对于来自问题（svmh2）的软边界svm的一个版本，通过增加额外的自由度得到，也就是说，我们使用边界δ=η/kwk，其中，η是我们希望最大化的一些正常数。为此，我们在目标函数中添加一个术语−kmη。我们有以下优化问题：

最小化



其中，km>0和ks>0是固定常数，可调整以确定η和正则化项的影响。

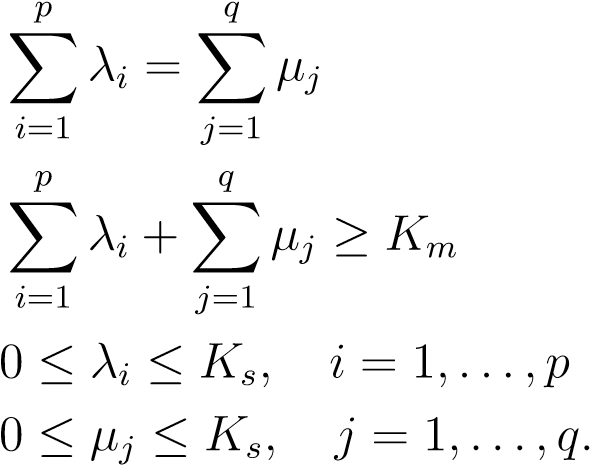
SVM问题的这一版本首先在scho–lkopf、smola、williamson和bartlett[143]中以ν-svc的名义进行了讨论，也用于sch–olkopf、platt、shawe–taylor和smola[142]。

为了解决这个问题，我们必须选择km和ks，以便

km≤最小2pks，2qks。

如第54.3节所示，双程序

最小化



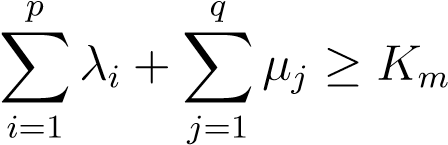
如果原问题有一个w=06的最优解，那么利用对偶间隙为零的事实，我们可以证明η≥0。因此，可以省略约束η≥0。

与前一种情况一样，w由

，

但b和η不是由对偶决定的。

如果我们降低约束η≥0，那么不等式



替换为公式

PQ

十倍

λi+μj=公里。I=1 J=1

方便地定义v>0，以便

km=（p+q）ksν，

那就是

，

因此，目标函数）由

，

k=（p+q）ks，所以km=kν，ks=k/（p+q）。

观察条件km≤min 2pks，2qks等于

.

由于我们通过用一个公共的正因子重新标定来获得一个等价的问题，因此可以方便地将ks归一化为

，

在这种情况下，km=ν。这种方法被称为nv-支持向量机。

在（svms20）的标准裕度假设下，存在一些i0，例如0<λi0<ks，一些j0，例如0<μj0<ks，并且由互补松弛条件=0和ξj0=0，因此我们有两个主动约束。

w>ui0−b=η，−w>vj0+b=η，

我们可以解b和η，得到

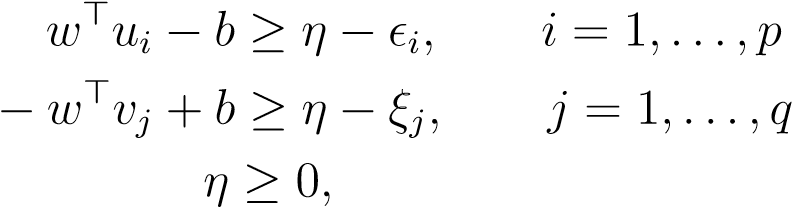
η=w>ui0−w>vj0。二

命题54.1给出了积分Ui的个数和未能达到边缘且最大有边缘的积分Vj的个数的上界。因此，如果uis和vjs不是线性可分离的，我们必须选择v，使2/（p+q）≤v≤min 2p/（p+q），2q/（p+q）方法成功。

我们还研究了确保某个点ui被正确分类或某个点vi被正确分类的条件，并且相应的约束是活动的（因此ui在边缘，resp）。VJ在边缘）。如果存在pf误分类点ui和qf误分类点vj，那么如果pf+qf≥3和2/（p+q）<（pf+qf）/（p+q），则上述属性保持不变；见54.2号提案。我们还表明，如果pf，qf≥2，如果2/（p+q）<4/（p+q），那么b和η可以在不参考标准裕度假设的情况下找到；见命题54.3。

（3）基本二次型软边-SVM问题（SVMS3）。这是问题（SVMS20）的版本，其中我们使用二次函数代替线性函数作为正则函数。优化问题是

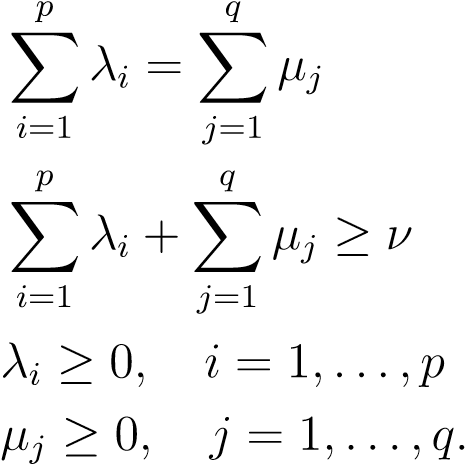
最小化



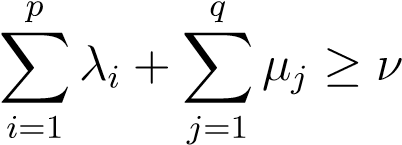
式中，ν和k是两个给定的正常数。如前所述，选择k=1/（p+q）比较方便。

在这种方法中，不再需要0，因为最优解满足这些条件。我们也可以省略约束η≥0，因为对于最优解，可以用对偶性表示η≥0。如第54.4节所示，对偶函数由下式给出：

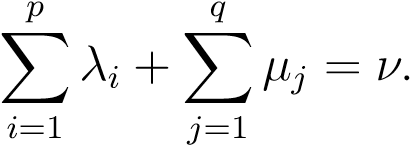
最小化



上述程序与为问题（SVMS20）而得到的程序相似，但矩阵x>x被矩阵x>x+（1/2k）ip+q所代替，该矩阵自k>0起为正定，且不再存在λi≤k和μj≤k的不等式。然而，这些限制意味着存在一些i0，例如λi0>0，而一些j0，例如μj0>0。如果约束η≥0下降，则不等式



替换为公式



我们从问题（SVMS20）中的λ和μ以及γ中获得w；即，

，

但对偶不确定b和η。然而，决定于

.

也因为约束

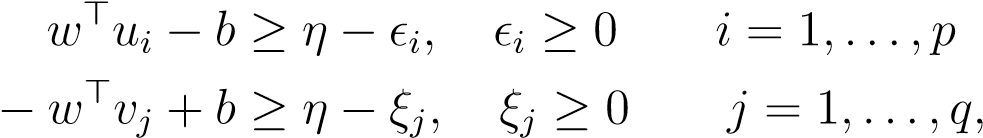
和

意味着存在一些i0，使得λi0>0，而一些j0，使得μj0>0，我们有

0和ξj0>0，这意味着至少两个点被错误分类，因此问题（SVMS3）只能在集合ui和vj不可线性分离时使用。我们可以使用与任何i0对应的主动约束来求解b和η，从而使λi0>0和任何j0，从而使μj0>0。这种方法不需要标准边际假设。

（4）软边ν-SVM问题（SVMS4）。这是通过在目标函数中添加术语（1/2）b2而得到的问题（SVMS20）的变化。结果表明，在将拉格朗日最小化的情况下，不仅可以求出双函数G，而且还可以求出B。我们还抑制了约束η≥0，这是多余的。优化问题是

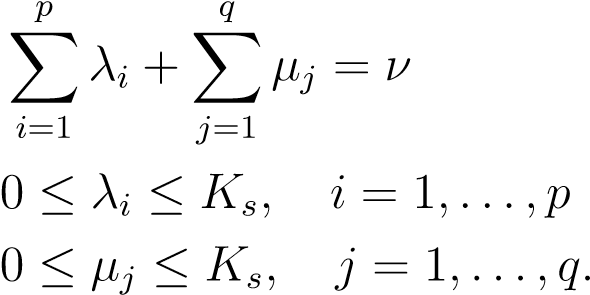
最小化



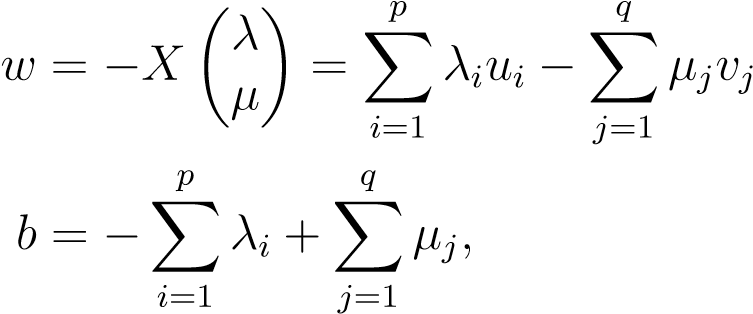
Ks=1/（P+Q）。

在第54.5节中显示，双变量由下式给出：

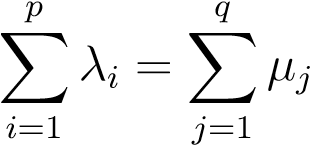
最小化



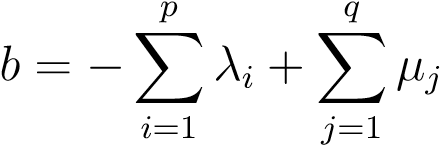
一旦得到λ和μ的溶液，我们就得到



但η不是由对偶决定的。注意约束



出现在程序对偶（SVMS20）中的已被交易为方程



确定b。这似乎是问题的优势（svms4）。

结果还表明，如果原始问题（SVMS4）有一个W=06的最优解，则η≥0。为了让原始人有一个解决方案，我们必须

ν≤1.

在（svms4）的标准裕度假设下，要么存在一些i0，例如0<λi0<ks，要么存在一些j0，例如0<μj0<ks，并且通过互补松弛条件=0，因此我们得出

w>ui0−b=η，或−w>vj0+b=η，

我们可以求出η。

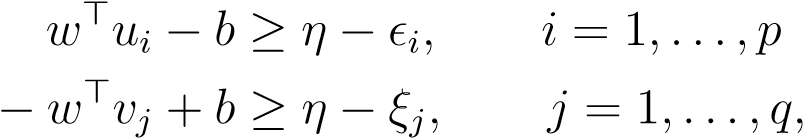
命题54.4给出了积分Ui的个数和未能达到边缘且最大有边缘的积分Vj的个数的上界。因此，如果uis和vjs不是线性可分离的，我们必须选择v，这样1/（p+q）≤v≤1，方法才能成功。

我们还研究了确保某个点ui被正确分类或某个点vi被正确分类的条件，并且相应的约束是活动的（因此ui在边缘，resp）。VJ在边缘）。如果存在pf错误分类点ui和qf错误分类点vj，则如果pf+qf≥2和1/（p+q）<（pf+qf）/（p+q），则上述属性保持不变。见54.5号提案；这比54.2号提案稍有改进。我们还表明，如果pf+qf≥2，并且如果1/（p+q）<3/（p+q），则无需标准裕度假设即可得出η；见命题54.6。这也比提议稍有改进

54.3。

（5）二次侧软边Ⅴ-SVM问题（SVMS5）。这是问题（SVMS3）的变体，其中我们将术语（1/2）b2添加到目标函数中。我们还降低了约束η≥0，这是多余的。我们有以下优化问题：

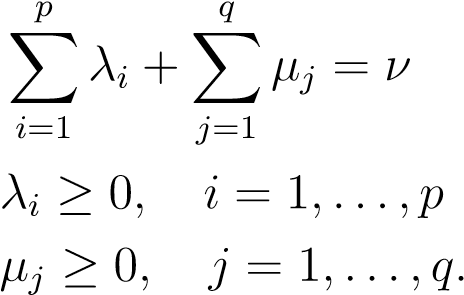
最小化



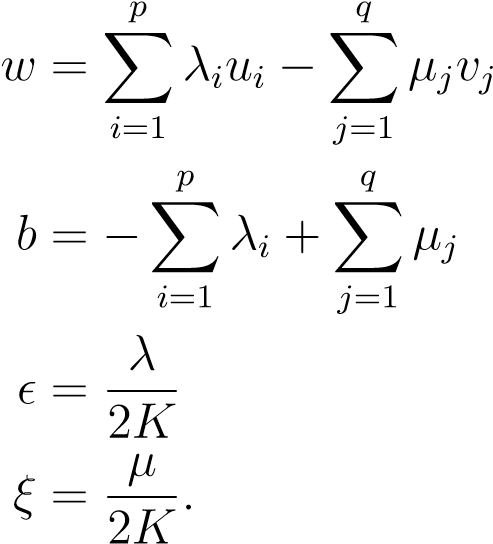
式中，ν和k是两个给定的正常数。如前所述，选择k=1/（p+q）比较方便。

在第54.6节中，程序对偶（SVMS5）由以下公式给出：

最小化



这一次，我们从λ和礹中得到和ξ：



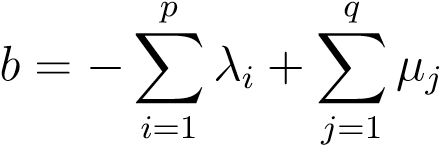
约束条件

PQ

十倍

λi=μj i=1 j=1

程序对偶（svms3）中出现的问题已被转换为方程。



确定B。这似乎是问题的优势（SVMS5）。约束条件

PQ

十倍

λi+μj=νi=1 j=1

意味着要么存在一些i0，比如λi0>0，要么存在一些j0，比如μj0>0，我们有0，这意味着至少有一个点被错误分类，所以只有当集合ui和vj不可线性分离时，才应使用问题（svms5）。我们可以使用与任何i0对应的主动约束来求解η，这样λi0>0或任何j0，这样μj0>0。利用对偶性可以证明，如果初等有一个w=06的最优解，则η≥0。

这些方法都有一个内核化的版本。

总之，从理论上看，问题（SVMS4）和（SVMS5）似乎比其他问题具有更多的优势，因为它们至少确定了w和b，但这有待于实验验证。

1860年第54章。软边界支持向量机

第十部分

附录

一千八百六十一

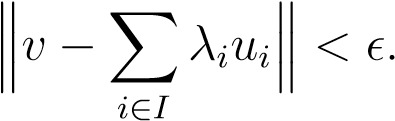
附录A

# 希尔伯特空间中的总正交族

## A.1总正交族（希尔伯特基），傅立叶系数

通过证明正交基的概念可以推广到希尔伯特空间，从而结束了希尔伯特空间的快速浏览。然而，有用的概念不是通常的基础概念，而是傅立叶级数概念的抽象概念。希尔伯特空间的每一个元素都是它的傅立叶级数的“和”。

定义A.1.对于希尔伯特空间e，非零向量的一个族（uk）k∈k是一个正交族iff，uk是对正交的，即hui，uji=0代表所有i=6 j（i，j∈k），而正交族iff hui，uji=δi，j代表所有i，j∈k，一个全正交族（或系统）或希尔伯特基是一个ort。在e中稠密的Hogonal族。这意味着对于每个v∈e，对于每个>0，都有一些有限子集i k和一些复数的族（λi）i∈i，这样



给定一个正交族（uk）k∈k，对于每一个v∈e，对于每一个k∈k，标量ck=hv，uki/kukk2称为v over（uk）k∈k的k次傅立叶系数。

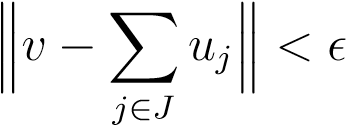
备注：hilbert基这个术语是误导性的，因为hilbert基（uk）k∈k不一定是代数意义上的基。事实上，一般来说，（uk）k∈k不跨越e。直观地说，它是利用具有无穷多非零系数的uk的线性组合来跨越e。从技术上讲，这是在极限方面实现的。为了避免代数意义上的基与希尔伯特基之间的混淆，一些作者将代数基称为哈默尔基，而将全正交族（或希尔伯特基）称为夏德基。

一千八百六十三

给定一个正交族（pλiui傅立叶级数的部分和）k∈k，对于任何有限子集，如果这些部分和收敛到k的a i，我们通常将表示为asi∈i p ckuk的形式极限的和称为pk∈k ckuk傅立叶级数。钾钾

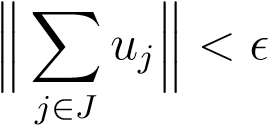
然而，我们必须理解这些数目！实际上，当k无序或不可数时，极限或和的概念还没有被定义。这可以按如下方式完成（有关更多详细信息，请参阅dixmier[52]）：

定义A.2.给定一个赋范向量空间e（即希尔伯特空间），对于任意一个非空索引集k，我们说e中向量的一个族（uk）k∈k可以用和v∈e iff求和，对于每个>0，都有k的一些有限子集i，这样，



对于具有i j k的每一个有限子集j，我们说族（uk）k∈k是可求和的，如果有一些v∈e，那么（uk）k∈k是可求和的。

Cauchy族iff对于每>0，都有k的有限子集i，这样，



对于i j=的k的每个有限子集j，

如果（uk）k∈k可与和v求和，我们通常将v表示为pk∈k uk。需要以下技术建议：

提案A.1.设e为完全赋范向量空间（例如，希尔伯特空间）。

1. 对于任何非空索引集k，一个族（uk）k∈k是可求和的，如果它是一个柯西族。
2. 给定一个非负实数Rk≥0的族（Rk）k∈k，如果有实数b>0

psuch that k∈k rk=pr，其中ri<rbis有限集的最小上界，对于k的每一个有限子集i，那么（rk）k∈kpis可求和i∈i ri（i k）。我爱我

证据。（1）如果（uk）k∈k是可和的，对于k的每个有限子集i，让

U=X英国

k∈k对于每>0，有k的有限子集i，这样



对于所有有限子集l，如果i l k，对于k的每个有限子集j，当i j k和i j是有限的，我们有

2和，

由于kui j−uik≤kui j−uk+ku−uik

而ui j−ui=uj，因为i j=∅，我们得到



这是（英国）K∈K成为柯西家族的条件。

相反，假设（uk）k∈k是柯西族。我们归纳地定义了e子集的一个递减序列（xn），每个直径至多为1/n，如下所示：对于n=1，因为（uk）k∈k是一个柯西族，k的一些有限子集j1使得

kujk<1/2

对于k的每一个有限子集j，j1 j=∅。我们选取一些具有上述性质的有限子集j1，我们让i1=j1和

x1=ui i1 i k，i有限。

对于n≥1，有一些k的有限子集jn+1，这样

kujk<1/（2n+2）

对于k的每一个有限子集j，jn+1 j=∅。我们选取了一些具有上述性质的有限子集jn+1，我们让+1=in jn+1和

xn+1=ui in+1 i k，i有限。

由于in in+1，很明显xn+1 xn适用于所有n≥1。我们需要证明每个xn的直径最多为1/n。因为jn的选择是这样的

kujk<1/（2n）

对于k的每一个有限子集j，jn j=∅，并且由于jn in，也确实

kujk<1/（2n）

对于k的每一个有限子集j，in j=（因为in j=和jn in意味着jn j=）。然后，对于每两个有限子集j，l，在j，l k中，我们得到

kuj−ink<1/（2n）和kul−ink<1/（2n），

由于kuj−ulk≤kuj−uink+kuin−ulk=kuj−ink+kul−ink，

我们得到

kuj−ulk<1/n，

这证明了δ（xn）≤1/n。现在，如果我们考虑闭集序列（xn），我们可以

仍然有xn+1 xn，根据命题47.4，δ（xn）=δ（xn）≤1/n，这意味着limn→∞δ（xn）=0，根据命题47.4，由单个元素u组成。我们声称u是族（u k）k∈k的和。

对于每一个>0，有一些n≥1这样，并且由于u∈xm对于所有m≥1，有一些k的有限子集j0这样在j0和

，

其中in是涉及xn定义的k的有限子集。然而，由于δ（xn）≤1/n，对于k的每个有限子集j，在j中，我们有

，

从那以后

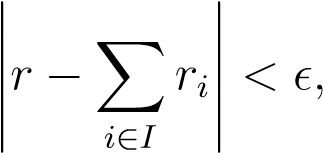
，

我们得到

对于k的每一个有限子集j，用in j证明u是族的和

（英国）K∈K.

（2）由于每一个有限和pi∈i ri都有一个统一的边界b，这些有限和的集合有一个最小上界r≤b。对于每一个>0，因为r是有限和pi∈i ri的最小上界（其中i是有限的，i k），有一些有限i k这样

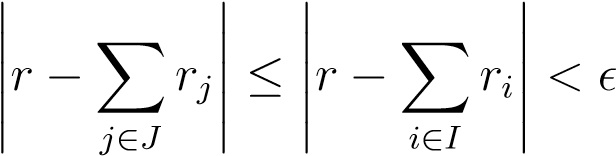


既然所有k∈k的Rk≥0，我们有

十倍

ri≤rj i∈i j∈j

每当我j，这表明



对于i j k这样的每一个有限子集j，证明（rk）k∈k可与和pk∈k rk=r求和。

注：可和性的概念意味着一个家族（uk）k∈k的和独立于k上的任何阶，在这个意义上，它是一种“交换可和性”。更准确地说，很容易看出，对于每个双射，如果k:k→k（直观地，是k的重新排序），k∈k是可求和的，如果k是可求和的，那么它们具有相同的和。

下面的命题给出了傅立叶系数的一些主要性质。除此之外，最多可数的傅立叶系数中的许多可能是非零的，并且傅立叶级数的部分和收敛。给定一个正交族（uk）k∈k，我们让uk=cuk，而puk:e→uk是e对uk的投影。

提案A.2.设e为希尔伯特空间，（uk）k∈k为e中的正交族，v为（uk）k∈k生成的子空间的闭包，其性质如下：

1. 对于每个v∈e，对于每个有限子集i k，我们有

，

式中，ck是v的傅里叶系数。

1. 对于每一个向量v∈e，如果（ck）k∈k是v的傅立叶系数，则下列条件是等效的：

（2a）v∈v

（2b）族（ckuk）k∈k是可和的，v=pk∈k ckuk。

（2c）族（ck 2）k∈k可求和，kvk2=pk∈k ck 2；

1. 族（ck 2）k∈k是可和的，我们得到了贝塞尔不等式：

.

因此，至多可数地说，许多CK可能是非零的。

（inckeuk）对某些向量k∈k形成一个柯西族，因此，傅立叶级数u=pk∈k ckuk。此外，u=pv（v）。pk∈k-ckuk收敛

证据。（1）让

ui=xciui

我爱我

对于k的任何有限子集i，我们声称v-ui与ui是正交的，对于每个i∈i。

\*+

hv−ui，uii=v−xcjuj，ui

J i

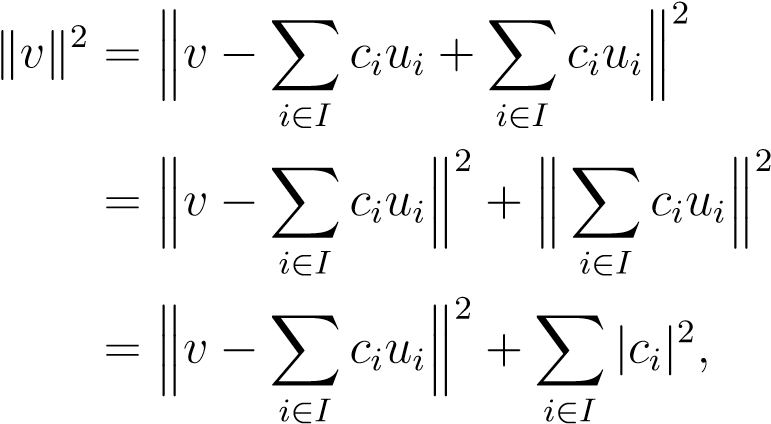
=hv，uii−xcj-huj，uii

J i

=hv，uii−cikuik2

=hv，uii−hv，uii=0，

由于huj，所有i=6 j和ci=hv，uii/kuik2的uii=0。因此，我们



因为UI是成对正交的，也就是说，

.

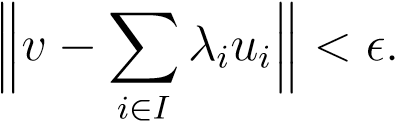
因此，

，

如要求。

（2）我们证明了影响链（a）⇒（b）⇒（c）⇒（a）。

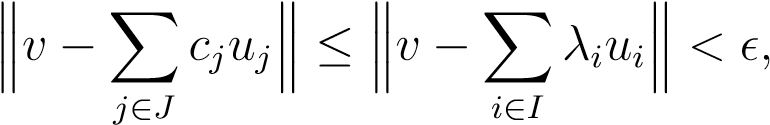
1. ？（b）：如果v∈v，由于v是（uk）k∈k所跨越的子空间的闭包，对于每个>0，存在k的一些有限子集i和复数的一些族（λi）i∈i，这样



现在，对于k的每一个有限子集j，比如i j，我们有

，

由于i j和uj（带有j∈j）与v−pj∈j cjuj是正交的，这表明



因此，家族（ckuk）k∈k可与和v求和，因此

.

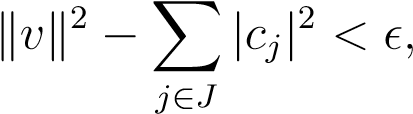
1. ？（c）：如果v=pk∈k-ckuk，那么对于每>0，存在k的有限子集i，这样



对于k的每一个有限子集j，这样i j，并且因为我们在（1）中证明了

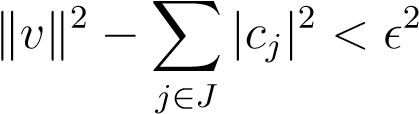
，

我们得到



证明（ck 2）k∈k可与和kvk2求和。

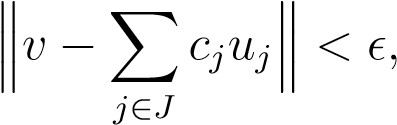
1. ？（a）：最后，如果（ck 2）k∈k可与和kvk2求和，则对于每个>0，都有k的一些有限子集i，这样



对于k的每一个有限子集j，这样i j，再次使用

，

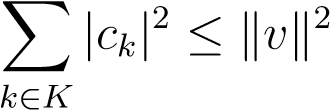
我们得到



证明了（ckuk）k∈k可与和pk∈k ckuk=v和v∈v求和。

|（3）2）k∈sincek可求和。对于k的每个有限子集i，贝塞尔不等式∈i ci 2≤kvk2，通过命题A.1，族

（CK）

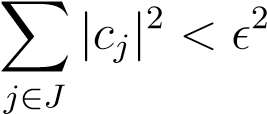


是不等式pi∈i ci 2≤kvk2（对于每个有限i k）的一个明显结果。现在，对于每一个自然数n≥1，如果kn是由所有ck组成的k的子集，那么ck≥1/n，kn中的元素数最多为

，

它是有限的，因此，ck的可数至多可以是非零的。

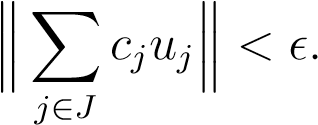
由于（c k 2）k∈k可与和c求和，对于每>0，存在k的一些有限子集i，因此



对于k的每一个有限子集j，i j=∅。自从

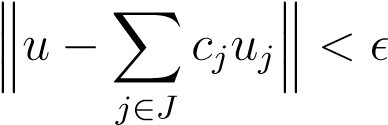
，

我们得到



（这证明（ckuk）k∈k是可和的，sinceckuk）k∈k是一个柯西族，通过命题A.1，它意味着这是完全的。因此，傅立叶级数pk∈k-ckuk是可和的，其和表示为u∈v。

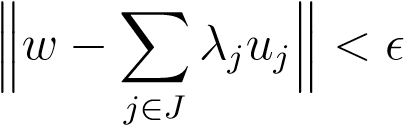
由于pk∈k-ckuk可与sum u求和，对于每>0，k中有一些有限子集i1，这样



对于k的每一个有限子集j，通过三角形不等式，对于k的每一个有限子集i，

.

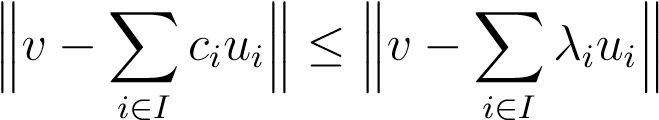
由（2）可知，每一个w∈v是其傅立叶级数pk∈kλkuk的和，并且对于每一个>0，存在k的一些有限子集i2，因此



对于k的每一个有限子集j，通过三角形不等式，对于k的每一个有限子集i，

.

假设i=i1 i2，因为我们在（2）中展示了



对于k的每个有限子集i，我们得到

，

因此



因为每大于0，我们就有

ku−vk≤kv−wk

对于所有w∈v，即kv−uk=d（v，v），用u∈v证明u=pv（v）。

## A.2希尔伯特空间` 2（k）和里兹费希尔定理

命题A.2建议观察序列（2）k∈k的空间是可和的。实际上，这些空间是希尔伯特空间，结果证明Everyzk）k∈k（其中zk∈c）这样

（ZK）

希尔伯特空间是同构的其中之一。这些空间是通常欧几里得范数下空间cn的无限维版本。

定义A.3.对于任何非空索引集k，空间2（k）是所有序列（zk）k∈k的集合，其中zk∈c，这样（zk 2）k∈k是可和的，即pk∈k zk 2<∞。

评论：

1. 当k是一组有限的基数n时，`2（k）与cn同构。
2. 当k=n时，空格'2（n）对应于第13.1节（第一卷）中实施例2的空格'2。在那个例子中，我们声称'2是一个隐士空间，实际上是一个希尔伯特空间。我们现在证明了任何索引集k的这个事实。

提案A.3.对于任何非空索引集k，空间2（k）是Hermitian积下的Hilbert空间。

.

由序列2（k）（zk）k∈k组成的子空间，使得zk=0，除了有限的多个k，是`

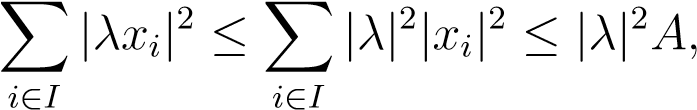
证据。首先，我们需要证明` 2（k）是一个向量空间。

（yk）k∈k在` 2（k）中。这意味着，（x xk 2）k k k和（yk yk 2）k k是可积的，从命题A.1的角度出发，相当于一些正界A和B的存在，使得πi i xi 2＜a和πi i yi 2＜b，对于k的每个有限子集i，证明了（xxk+yk\* 2）k k是可积的，这是足够的。证明了在k的每个有限子集i中都存在一些πi i i Xi＋Yi＞2＜C的C＞0，但平行四边形不等式意味着

，

对于k的每一个有限子集i，我们用命题A.1得出结论。同样，对于每个

λ∈c，



且（λkxk）k∈k可求和。因此，`2（k）是一个向量空间。

通过柯西-施瓦兹不等式，

，

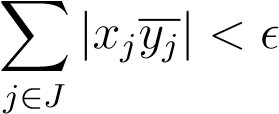
对于k的每一个有限子集i，这里我们使用

4c d≤（c+d）2，

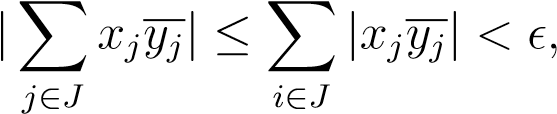
相当于

（c-d）2≥0。

根据命题A.1，（xkyk）k∈k是可和的。习惯用语是（xkyk）k∈k是绝对可和的。然而，这是一个标准的事实，这意味着（xkyk）k∈k是可和的（对于每>0，有一些k的有限子集i，这样

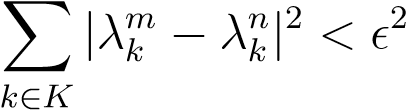


对于k的每一个有限子集j，i j=∅，因此



证明（xkyk）k∈k是柯西族，因而可求和）。我们还必须证明2（k）是完整的。

考虑一个序列（（λn k）k∈k）n≥1的序列（λnk）k∈k∈2（k），并假定它是一个柯西序列。这意味着，对于每>0，有一些n≥1，这样

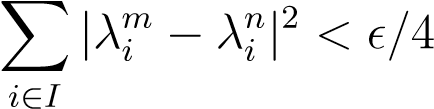


对于所有m，n≥n。对于每个固定k∈k，这意味着

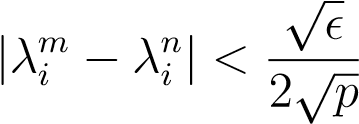


对于所有m，n≥n，这表明（是c中的一个柯西序列，由于c是完整的，所以序列（λn k）n≥1有一个极限λk∈c，我们声称（λk）k∈k∈2（k），这是（（λnk）k∈k）n≥1的极限。

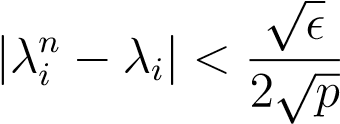
在任何>0的情况下（（λn k）k∈k）n≥1是一个柯西序列，这意味着存在一些n≥1，因此对于k的每个有限子集i，我们可以



对于所有m，n≥n。设p=i。然后，



对于每一个i∈i，由于λi是（λni）n≥1的极限，我们可以找到一些足够大的n，以便



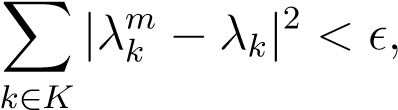
从那以后

，

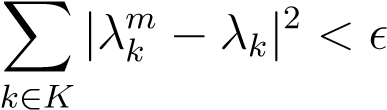
我们得到

因此，

对于所有的m≥n，由于上述方法适用于k的每一个有限子集i，通过命题A.1，我们得到



向量空间和（对于所有m≥n。这证明（λk）k∈k∈λkm−λk）k∈≥k∈2（k）对于所有m∈≥n2（，和sincek）。然而，`2（k）是m`2（k），对于所有m 1，我们得到（λk）k∈k`



对于所有m≥n，表示序列（收敛于（λk）k∈k∈2（k）。由序列（2（k）组成的子空间留作一个简单的练习。zk）k∈k使得zk=0，除了有限的多个k之外，对于有限的manyremark，子空间是一个稠密的suspace‘：由所有序列组成的子空间（k，提供了一个在zk中不封闭的子空间的例子）。k∈k使得zk=0，除了可能的'2（k）。实际上，这个空间严格包含在'2（k）中，因为'2（k）中有非空元素的可数序列（为什么？）.

在证明每个希尔伯特空间都同构于某些2（k）之前，我们只需要两个命题。

提案A.4.设e为希尔伯特空间，（uk）k∈k为e中的正交族，其性质如下：

1. 对于每个族（λk）k∈k∈2（k），该族（λkuk）k∈k是可和的。此外，v=pk∈kλkuk是唯一的向量，这样，所有k∈k的ck=λk，其中，ck是

傅立叶系数。

1. 对于任意两个族（λk）k∈k∈2（k）和（μk）k∈k∈2（k），如果v=pk∈kλkuk和w=pk∈kμkuk，我们得到以下方程，也称为parseval恒等式：

.

命题A.2（3）中给出的证明适用于家庭（即（（1）可求和的事实）。设（λk）k∈k∈2（k）表示V=λk 2）（λk（而不是2）k∈k是可求和的。证明（ck 2）k∈hk）和yieldsi k k2。

k k∈k p k k∈k∈kλk uk，回忆一下ck=v，uk/uk

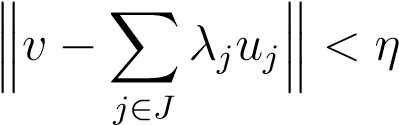
选取一些k∈k。由于h−，−i是连续的，对于每>0，有一些η>0，这样



无论何时

kv−wk<η。

然而，由于对于每个η>0，存在k的有限子集i，因此



对于k的每一个有限子集j，如i j，我们可以选取j=i k，并让w=pj∈jλjuj，我们得到

.

然而，

hv，uki=ckkukk2和，

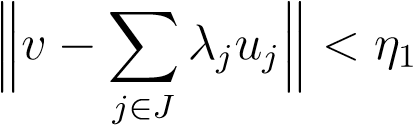
因此，上面证明了对于每>0，那么，ck=λk。

（2）由于H−、−I是连续的，对于每>0，有一些η1>0和η2>0，例如

那个

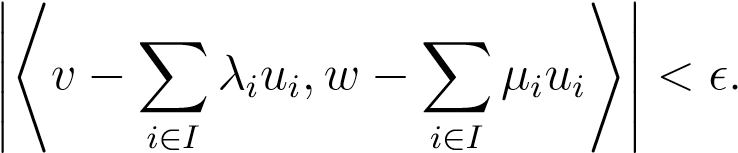


当Kxk<η1和Kyk<η2时。由于v=pk∈kλkuk和w=pk∈kμkuk，k中存在一些有限子集i1，因此



对于k的每一个有限子集j，比如i1 j，还有k的一些有限子集i2，这样

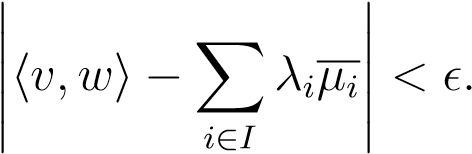
对于k的每一个有限子集j，如果I2 j.让i=I1 I2，我们得到



此外，



由于ui与所有i∈i的v−pi∈iλiui和w−pi∈iμiui是正交的，这证明了对于每个>0，都存在k的一些有限子集i，因此



从命题A.3我们已经知道（λkµk）k∈k是可和的，并且由于>0是任意的，我们得到

.

下一个命题陈述了表征希尔伯特基（总正交族）的性质。

提案A.5.设e为希尔伯特空间，设（uk）k∈k为e中的正交族。

以下属性等效：

1. 家族（uk）k∈k是一个完全正交的家族。
2. 对于每一个向量v∈e，如果（ck）kp∈k是ckuk.v的傅立叶系数，那么族

（ckuk）k∈k是可和的，v=k∈k

1. 对于每个向量v∈e，我们都有parseval恒等式：

.

1. 对于每一个向量u∈e，如果hu，uki=0，对于所有k∈k，则u=0。

证据。（1）、（2）和（3）的等价性是提案A.2和提案A.4的直接结果。

（4）如果（u k）k∈k是一个完全正交的族，而hu，uki=0代表所有k∈k，因为u=pk∈k-ckuk，其中ck=hu，uki/kukk2，我们得到k=0代表所有k∈k，u=0。

相反地，假设（u k）k∈k的闭包v不同于e，那么，根据命题47.7，我们得到e=v v，其中v是v的正交补，v是非平凡的，因为v=6e，因此存在一些非空向量u∈v。但是，u与v中的每一个向量都是正交的，特别是，

胡，Uki=0

对于所有k∈k，假设它意味着u=0，与u=6的事实相矛盾。

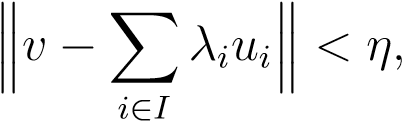
0。

评论：

1. 如果e是希尔伯特空间，并且（u k）k∈k是e中的一个完全正交族，那么基于h−i的连续性，有一个更简单的论证证明u=0，如果hu，uki=0，对于所有k∈k。论证是证明假设意味着h v，ui=0，对于所有v∈e。因为h−，−i是正的。我确定，这意味着u=0。通过h−、−i的连续性，对于每个>0，都有一些η>0，这样对于k的每个有限子集i，对于每个族（λi）i∈i，对于每个v∈e，

无论何时

由于（uk）k∈k在e中是稠密的，对于每个v∈e，都有k的有限子集i和一些族（λi）i∈i，因此



由于假设，=0，我们得到



由于此值适用于每>0，我们必须有hv，ui=0

1. 如果v是e的任何非空子集，则可以使用前面注释中使用的参数来证明v是闭合的（即使v不是），并且v是v的闭合。

现在我们将证明每个希尔伯特空间都有一些希尔伯特基。这需要使用集合论中的一个基本定理，即我们很快回顾的佐恩引理。

对于任何具有部分排序≤的集合x，回想一下，如果x的非空子集c是完全有序的（即，对于所有x，y∈c，x≤y或y≤x），那么它就是链。x的一个非空子集y是有界的，如果有一些b∈x，使得y≤b代表所有y∈y。对于每一个x，m∈x是极大的iff，m≤x意味着x=m，现在我们可以描述Zorn的引理。有关详细信息，请参见Rudin[136]、Lang[106]或Artin[7]。

提案A.6.给定任意非空部分序集x，如果x中的每个（非空）链都有界，那么x有一些极大元素。

我们现在可以证明希尔伯特基的存在。我们在族（uk）k∈k上定义了一个偏序：对于任意两个族（uk）k∈k1和（vk）k∈k2，我们说

（UK）K∈K1≤（VK）K∈K2

iff k1 k2和uk=vk表示所有k∈k1。这显然是部分订单。

提案A.7.让我们成为希尔伯特空间。在E中任意一个正交族（uk）k∈k，都有一个完全正交族（ul）l∈l包含（uk）k∈k。

证据。考虑到所有正交族的集合s大于或等于b=（uk）k∈k，我们认为s中的每个链都是有界的。实际上，如果c=（c l）l∈l是s中的链，其中cl=（uk，l）k∈kl，则为联合族。

（uk）k∈sl∈l k l，其中uk=uk，l每当k∈k l，

显然是c的上界，并立即证明它是正交族。在Zorn引理A.6中，有一个极大族（u l）l∈l包含（u k）k∈k，如果（ul）l∈l在e中不稠密，则其闭包v严格包含在e中，根据命题47.7，由于v=6e，v的正交补码v是非平凡的，因此存在一些非零向量u。V。然后，u对（u l）l∈l中的每一个向量都是正交的，我们可以把u加到这个族中，从而形成严格大于（ul）l∈l的正交族，这与（ul）l∈l的最大性相矛盾，因此，（ul）l∈l在e中是稠密的，因此它是希尔伯特基。

备注：可以证明希尔伯特空间e的所有希尔伯特基都具有相同基数的索引集k。有关证据，请参见Bourbaki[27]。

定义A.4.如果希尔伯特基是可数的，那么希尔伯特空间e是可分离的。

最后证明了对于某些合适的K，每个希尔伯特空间都是同构的。

定理A.8。（Riesz Fischer）对于每个希尔伯特空间e，都有一些非空集k

使e与希尔伯特空间'2（k）同构。更具体地说，对于任何希尔伯特基（uk）k∈k of e，映射f：`2（k）→e和g:e→`2（k）定义如下：

f（（λk）k∈k）=xλkuk和，

钾钾

是双射线性等轴测，如g f=id和f g=id。

证据。根据命题A.4（1），图F定义得很好，并且它明显是线性的。通过命题A.2（3），图G被很好地定义，而且它也是清晰的线性的。根据提案A.2（2b），我们有

f（g（u））=u=x克，

钾钾

根据提案A.4（1），我们有

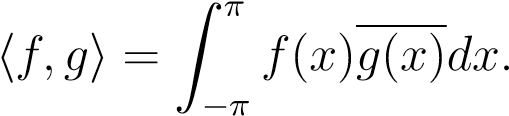
g（f（（λk）k∈k））=（λk）k∈k，

因此，g f=id和f g=id。根据命题A.4（2），线性图g是一个等距图。因此，F是一个线性双射体，在` 2（k）和e之间有一个等距线，反G。

注：图G:E→'2（k）的主客观性称为Riesz-Fischer定理。

在做了所有这些艰苦的工作之后，我们勾画出这些结果如何应用于傅立叶级数。我们再次向读者推荐鲁丁（Rudin）或朗（Lang）来进行全面的阐述。

设c（t）表示周期2π的所有周期连续函数f：[−π，π]→c的集合。存在一个含c（t）的希尔伯特空间l2（t），使得c（t）在l2（t）中稠密，其内积由



希尔伯特空间l2（t）是勒贝格平方可积周期函数（周期2π）的空间。

结果表明，在l2（t）中，族（eikx）k∈z是一个完全正交的族，因为它在c（t）中已经很密集（例如，参见rudin[136]）。然后，Riesz-Fischer定理指出，对于每个族（ck）k∈z的复数，这样

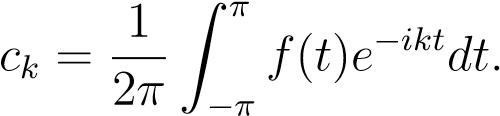
，

有一个唯一的函数f∈l2（t），使得f等于它的傅立叶级数。

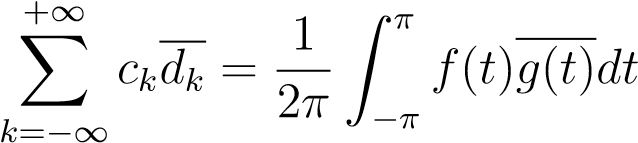
f（x）=xckeikx，

K z

其中f的傅里叶系数ck由公式给出



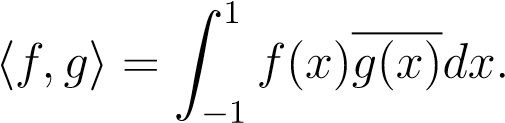
帕西瓦尔定理说



对于所有f，g∈l2（t），其中，ck和dk是f和g的傅立叶系数。

因此，希尔伯特空间l2（t）和'2（z）之间存在同构，这就是傅立叶系数“起作用”的深层原因。定理A.8表明函数f∈l2（t）的傅立叶级数pk∈z-ckeikx在l2意义上收敛到f，即在均方意义上。这并不一定意味着傅立叶级数收敛到f点！这是一个微妙的问题，关于这个主题，读者可以参考lang[108，109]或Schwartz[148，149]。

我们还可以考虑连续函数f：[-1,1]→c的集合c（[-1,1]）。存在一个包含c（[-1,1]）的希尔伯特空间l2（[-1,1]），其中c（[-1,1]）在l2（[-1,1]）中密集，其内积由下式给出。



希尔伯特空间l2（[-1,1]）是勒贝格平方可积函数在[-1,1]上的空间。第11.2节（第1卷第11章）实施例5中定义的勒让德多项式pn（x）构成了l2的希尔伯特基（[-1,1]）。回想一下，如果我们让fn成为函数

Fn（x）=（x2-1）N，

Pn（x）定义如下：

p0（x）=1，和，

式中，Fn的第n个导数，超前系数的原因是

PN（1）=1。这可以用很多努力来证明

.

附录B

# 佐恩引理；一些应用

## B.1佐恩引理的陈述

佐恩引理是选择公理的一种特别有用的形式，特别是在代数应用中。想要进一步了解Zorn引理及其在代数中的应用的读者，可以参考lang[106]、附录2、？2（第878-884页）和第三章、？5（第139-140页）或Artin[7]、附录1（第588-589页）。对于Zorn引理的逻辑分支及其与选择公理的等价性，我们应该参考Schwartz[146]，（第1卷），第一章，？，6，或集合理论的文本，如Enderton[57]，Suppes[168]或Kuratowski和Mostowski[105]。

给定一个集合，s，一个偏序，≤，on s是s上的二元关系（即≤s×s），即

1. 反身性，即x≤x，对于所有x∈s，
2. 可传递的，即，如果x≤y和y≤z，那么x≤z，对于所有x，y，z∈s，和
3. 反对称性，即，如果x≤y和y≤x，则x=y，对于所有x，y∈s。

一对（s，≤），其中≤是s上的部分序，称为部分序集或偏序集。

元素在一个位置集内，（x，y∈cs，，或≤），一个子集，x≤yif有某种力，of≤xs。空的那套简直就是一条链子。一个子集，是完全有序的b∈ssso，这是非空的。aor ax≤chainb，对于每对x∈p.observep的所有最大元素ofif，

s的（空或不空）是有界的

如果且仅当

p是一个元素，m∈p，因此m≤x意味着m=x，对于所有x∈p，zorn的引理可以表述如下：

引理B.1。给定一个部分有序集，（s，≤），如果每个链都有界，则s有一个极大元素。

证据。参见Schwartz[146]、Enderton[57]、Suppes[168]或Kuratowski和Mostowski中的任何一个。

〔105〕。

一千八百八十一

附录B.Zorn引理；一些应用

注：正如我们所注意到的，zorn引理的假设意味着s是非空的（因为空集必须是有界的）。一个部分有序的集合，使得每个链都有界，有时称为归纳集。

我们现在给出了Zorn引理的一些应用。

## B.2向量空间中基存在的证明

利用Zorn的引理，我们可以证明定理3.5适用于任意向量空间，而不仅仅适用于有限生成的向量空间，如第3章所述。

定理B.2.给定任一族，s=（ui）i∈i，生成向量空间e和任意线性无关的子族，l=（uj）j∈j，其中s（其中j i）有e的基，b，这样

L B S.

证据。考虑线性独立族的集合l，b，这样l b s。由于l∈l，这个集合是非空的。我们声称l是归纳的。考虑L中线性独立族b l的任意链（bl）l∈∧，并考虑b=sl∈∧∧bl。对于某些指数集h，b族的形式为b=（vh）h∈h，它必须是线性独立的。事实上，如果这不是真的，就会有一个有限支持的标量族（λh）h∈h，因此

X

λhvh=0，h∈h

其中并非所有的λh都为零。由于b=sl∈∧bl且只有有限多个λh为非零，因此∧有一个有限子集f，因此vh∈bfh iffλh=06。但是（bl）l∈∧是一条链，如果我们让f=max fh fh∈f，那么vh∈bf，对于所有的vh，其中λh 6=0。因此，

X

λhvh=0

H·H

将是一个非平凡的线性依赖向量从高炉，一个矛盾。因此，b∈l，由于b显然是bl的上界，我们证明了l是归纳的。利用Zorn的引理（LemmaB.1），集合L有一些极大的元素，比如b=（uh）h∈h，其余的证明与定理3.5的证明是一样的，但为了方便读者，我们重复它。我们认为b产生e。实际上，如果b不产生e，那么在b中有一些上∈s不是向量的线性组合（因为s产生e），带有p/∈h。然后，通过引理3.4，族b0=（uh）h∈h p是线性独立的，并且由于l b b0 s，这个逆ICT是b的最大值。因此，b是e的基础，因此l b s。

Zorn引理的另一个重要应用是极大理想的存在。

B.3.最大适理想的存在1883

## B.3包含给定适当理想的最大理想的存在性

设A为具有单位元的交换环。如果A=6 A，则A中的理想A是适当的理想。以下定理成立：

定理B.3.对于任何合适的理想，a a，有一个最大理想b，包含a。

证据。让我是包含a的所有适当理想的集合，b。集合i是非空的，因为a∈i。我们声称i是归纳的。考虑A中理想a i的任意链（ai）i∈i，可以很容易地检查b=si∈ai是一个理想。此外，B是一个合适的理想，因为

我

否则，单位元1将属于b=a，因此，对于某些i，我们将有1∈ai，这意味着ai=a，一个矛盾。另外，B显然是所有人工智能的上界，根据Zorn的引理（引理B.1），集I有一个极大的元素，比如B，B是一个包含A的极大理想。

1884附录B.Zorn的引理；一些应用

# 参考文献

1. 拉尔夫·亚伯拉罕和杰罗尔德·E·马斯登。力学基础。艾迪生·韦斯利，第二版，1978年。
2. 拉尔斯诉阿弗斯和利奥·萨里奥。黎曼曲面。普林斯顿数学。系列，2号。普林斯顿大学出版社，1960年。
3. 乔治·E·安德鲁斯、理查德·阿斯基和兰扬·罗伊。特殊功能。剑桥大学出版社，第一版，2000年。
4. 汤姆·阿波斯托。分析。艾迪生·韦斯利，第二版，1974年。
5. V.I.阿诺德。经典力学的数学方法。GTM第102号。Springer Verlag，第二版，1989年。
6. 埃米尔·阿廷。几何代数。Wiley Interscience，第一版，1957年。
7. 迈克尔·阿廷。代数。普伦蒂斯·霍尔，第一版，1991年。
8. M.F.Atiyah和I.G.MacDonald。交换代数导论。艾迪生·韦斯利，第三版，1969年。
9. A.阿维斯。计算差异。马森，第一版，1991年。
10. 谢尔顿·阿克斯勒。线性代数做得很好。数学本科课本。Springer Verlag，第二版，2004年。
11. 马塞尔·伯杰。G’EOM’ETRIE 1.内森，1990年。英语版：几何学1，大学出版社，斯普林格出版社。
12. 马塞尔·伯杰。G’EOM’ETRIE 2.内森，1990年。英语版：几何学2，大学版，斯普林格出版社。
13. MarcelBerger和Bernard Gostiaux。G’eom’etrie differ’erentielle:vari’et’es，courbes et surfaces.收集数学模型。Puf，第二版，1992年。英文版：微分几何，流形，曲线和表面，gtm第115号，Springer Verlag。
14. 罗尔夫·伯恩特辛几何导论。数学研究生，第26卷。AMS，第一版，2001年。

一千八百八十五

1. J.E.伯坦。alg'ebre lin'eaire et g'eom'etrie经典。马森，第一版，1981年。
2. Dimitri P.Bertsekas。凸优化理论。雅典娜科学，第一版，2009年。
3. Dimitri P.Bertsekas。凸优化算法。雅典娜科学，第一版，2015年。
4. Dimitri P.Bertsekas。非线性规划。雅典娜科学，第三版，2016年。
5. Dimitri P.Bertsekas、Angelina Nedi'c和Asuman E.Ozdaglar。凸分析与优化。雅典娜科学，第一版，2003年。
6. Dimitri P.Bertsekas和John N.Tsitiklis。并行和分布式计算：数值方法。雅典娜科学，1997年第一版。
7. Dimitris Bertsimas和John N.Tsitiklis。线性优化导论。雅典娜科学，第三版，1997年。
8. A.Beutelspacher和U.Rosenbaum。射影几何。剑桥大学出版社，第一版，1998年。
9. 克里斯托弗·M·毕晓普。模式识别和机器学习。信息科学与统计学。斯普林格，第一版，2006年。
10. 尼古拉斯·波巴基。代数，第9章。数学符号。赫尔曼，1968年。
11. 尼古拉斯·波巴基。代数，第1-3章。数学符号。赫尔曼，1970年。
12. 尼古拉斯·波巴基。代数，第4-7章。数学符号。马森，1981年。
13. 尼古拉斯·波巴基。ESPACES矢量拓扑。数学符号。马森，1981年。
14. 史蒂芬·博伊德、尼尔·帕里克、埃里克·朱、博贾·佩莱托和乔纳森·埃克斯坦。利用乘法器的交替方向法进行分布优化和统计学习。机器学习的基础和趋势，3（1）：1–122，2010年。
15. 史蒂芬·博伊德和列文·范登伯格。凸优化。剑桥大学出版社，第一版，2004年。
16. 格伦·布雷登。拓扑和几何。GTM第139号。Springer Verlag，第一版，1993年。
17. 海姆·布雷兹。泛函分析、索波列夫空间和偏微分方程。

大学文本。Springer Verlag，第一版，2011年。

1. G.Cagnac、E.Ramis和J.Commeu。Math’Ematiques Sp’Eciales，第3卷，G’eom’Etrie。马森，1965年。
2. 伊莱·卡坦。旋转器理论。多佛，第一版，1966年。
3. 亨利·卡坦。计算差异的过程。收集M’Ethodes。赫尔曼，1990年。
4. 亨利·卡坦。微分形式。多佛，第一版，2006年。
5. 张志忠和林志仁。培训霏-支持向量分类器：理论和算法。神经计算，13:2119–2147，2001。
6. 克劳德·切瓦利。旋量和克利福德代数的代数理论。作品集，第二卷。斯普林格，第一版，1997年。
7. Yvonne Choquet Bruhat、C'ecile Dewitt Morette和Margaret Dillard Bleick。分析，流形和物理，第一部分：基础。北荷兰，1982年第一版。
8. 范立忠。谱图理论，第92卷区域会议系列数学。AMS，第一版，1997年。
9. Vasek Chvatal公司。线性规划。W.H.弗里曼，第一版，1983年。
10. P.G.Ciarlet。数值矩阵分析与优化导论。剑桥大学出版社，第一版，1989年。法文版：Masson，1994。
11. Timoth'ee Cour和Jianbo Shi。用谱松弛法求解马尔可夫随机场。在玛丽塔·梅拉和沈小童，编辑，人工智能和统计。人工智能与统计学会，2007年。
12. H.S.M.考克斯特。非欧几里得几何。多伦多大学出版社，第一版，1942年。
13. H.S.M.考克斯特。几何概论。威利，第二版，1989年。
14. H.S.M.考克斯特。真正的投影平面。Springer Verlag，第三版，1993年。
15. H.S.M.考克斯特。射影几何。Springer Verlag，第二版，1994年。
16. 加贝。乘数法在变分不等式中的应用。数学与应用研究，15（c）：299-3311983。
17. 加斯顿·达布克斯。G’eom’etrie分析原理。高蒂尔·维拉斯，第一版，1917年。
18. 詹姆斯·W·德梅尔。应用数值线性代数。暹罗出版物，第一版，1997年。
19. Jean Dieudonn'e.Alg'ebre Lin'Eaire等人赫尔曼，第二版，1965年。
20. Jean Dieudonn'e.Sur Les Groupes Classiques公司。赫尔曼，第三版，1967年。
21. 雅克·迪克斯米尔。一般拓扑。UTM。Springer Verlag，第一版，1984年。
22. 曼弗雷多·P·多·卡莫。曲线和曲面的微分几何。普伦蒂斯·霍尔，1976年。
23. 曼弗雷多·P·多·卡莫。黎曼几何。Birkha–用户，第二版，1992年。
24. David S.Dummit和Richard M.Foote。抽象代数。威利，第二版，1999年。
25. 杰拉尔德·A·埃德加。度量、拓扑和分形几何。数学本科课本。Springer Verlag，第一版，1992年。
26. 赫伯特B.恩德斯顿。集合论的要素。学术出版社，1997年。
27. 查尔斯·L·爱泼斯坦。医学成像数学导论。暹罗，第二版，2007年。
28. 杰拉尔德·法林。CAGD的曲线和曲面。学术出版社，第四版，1998年。
29. 奥利维尔·法格拉斯。三维计算机视觉，几何视点。麻省理工出版社，1996年第一版。
30. 杰拉德·费舍尔。数学模型，评论。vieweg&sohn，第一版，1986年。[62]杰德·费舍尔。数学模型。vieweg&sohn，第一版，1986年。
31. 杰拉德·费舍尔。平面代数曲线。学生数学图书馆。AMS，第一版，2001年。
32. 詹姆斯·福利、安德烈·范丹、史蒂文·费纳和约翰·休斯。计算机图形。原则和实践。艾迪生·韦斯利，第二版，1993年。
33. 大卫·A·福塞思和让·庞斯。计算机视觉：现代方法。普伦蒂斯

霍尔，第一版，2002年。

1. 让·菲涅尔。现代人的思想。赫尔曼，第一版，1998年。
2. 威廉·富尔顿。代数曲线。高级书籍经典。艾迪生·韦斯利，第一版，1989年。
3. 威廉·富尔顿。代数拓扑学，第一门课程。GTM第153号。Springer Verlag，第一版，1995年。
4. 威廉·富尔顿和乔·哈里斯。代表理论，第一门课程。GTM第129号。Springer Verlag，第一版，1991年。
5. 让·加利尔。无符号图和有符号图的谱图理论。图形聚类的应用：一项调查。技术报告，宾夕法尼亚大学，2019年。

http://www.cis.upenn.edu/jean/spectral-graph-notes.pdf.

1. 让·H·加利尔。几何建模中的曲线和曲面：理论和算法。摩根考夫曼，1999年。
2. 让·H·加利尔。离散数学。大学文本。Springer Verlag，第一版，2011年。
3. 让·H·加利尔。用于计算机科学和工程的几何方法和应用。TAM，第38卷。斯普林格，第二版，2011年。
4. 让·H·加利尔。关于凸集、多面体、多面体、组合拓扑、Voronoi图和Delaunay三角测量的注释。技术报告，宾夕法尼亚大学独联体系，费城，PA 19104，2016。正在准备。
5. Walter Gander、Gene H.Golub和Urs von Matt。约束特征值问题。线性代数及其应用，114/115:815–839，1989。
6. 罗杰·戈登特。古斯·德·阿尔格。赫尔曼，第一版，1963年。
7. Chris Godsil和Gordon Royle。代数图理论。GTM编号207。斯普林格

Verlag，第一版，2001年。

1. 吉恩·H·戈卢布。一些修正的特征值问题。《暹罗评论》，15（2）：318–334，1973年。
2. Gene H.Golub和Frank Uhlig。QR算法：50年后，约翰·弗朗西斯和维拉·库布兰诺夫斯卡娅提出了QR算法的起源及其随后的发展。IMA数值分析杂志，29:467–485，2009年。
3. H.Golub，Gene和F.Van Loan，查尔斯。矩阵计算。约翰霍普金斯大学出版社，1996年第三版。
4. A.灰色。现代曲线和曲面的微分几何。CRC出版社，第二版，1997年。
5. 唐纳德·T·格林伍德。动力学原理。普伦蒂斯·霍尔，第二版，1988年。
6. 拉里·C·格罗夫。经典群和几何代数。数学研究生，第39卷。AMS，第一版，2002年。
7. 雅克·哈达马德。Le cons de g eom etrie el ementaire。I G’eom’etrie飞机。阿尔芒·科林，第十三版，1947年。
8. 雅克·哈达马德。Le cons de g eom etrie el ementaire。II G’EOM’ETRIE DANS L’ESPACE公司。阿尔芒·科林，第八版，1949年。
9. 约瑟夫·哈里斯。代数几何，第一门课程。GTM第133号。Springer Verlag，第一版，1992年。
10. 特雷弗·黑斯迪、罗伯特·提比西拉尼和杰罗姆·弗里德曼。统计学习的要素：数据挖掘、推理和预测。斯普林格，第二版，2009年。
11. 特雷弗·黑斯迪、罗伯特·提比西拉尼和马丁·温赖特。稀疏统计学习。套索和概括。CRC出版社，第一版，2015年。
12. 希古德尔·赫尔加森。分组和几何分析。积分几何、不变微分算子和球面函数。MSM，第83卷。AMS，第一版，2000年。
13. D.Hilbert和S.Cohn Vossen。几何学和想象。切尔西出版公司，1952年。
14. 莫里斯·W·赫斯和斯蒂芬·斯梅尔。微分方程、动力学系统和线性代数。学术出版社，第一版，1974年。
15. 罗杰A霍恩和查尔斯R约翰逊。矩阵分析。剑桥大学出版社，第一版，1990年。
16. 罗杰A霍恩和查尔斯R约翰逊。矩阵分析中的主题。剑桥大学出版社，第一版，1994年。
17. 克劳斯·胡雷克。初等代数几何。学生数学图书馆。AMS，第一版，2003年。
18. 内森·雅各布森。基础代数II。弗里曼，第一版，1980年。
19. 内森·雅各布森。基础代数I.弗里曼，第二版，1985年。
20. Ramesh Jain、Rangachar Katsuri和Brian G.Schunck。机器视觉。麦格劳-

希尔，第一版，1995年。

1. 朱尔根·乔斯特。黎曼几何与几何分析。大学文本。Springer Verlag，第四版，2005年。
2. Hoffman Kenneth和Kunze Ray。线性代数。普伦蒂斯·霍尔，第二版，1971年。
3. D.Kincaid和W.Cheney。数值分析。布鲁克斯/科尔出版社，第二版，1996年。
4. 菲利克斯·克莱恩。几何学，几何学。1927年，第一版。
5. 安东尼·W·纳普。除了引言之外，还可以分组说谎。数学进展，第140卷。Birkha–用户，第二版，2002年。
6. A.N.Kolmogorov和S.V.Fomin。介绍性真实分析。多佛，第一版，1975年。
7. 埃尔温·克雷塞格。微分几何。多佛，第一版，1991年。
8. K.Kuratowski和A.Mostowski。集合论。逻辑研究，第86卷。埃尔塞维尔，1976年。
9. 谢尔盖·朗。代数。艾迪生·韦斯利，第三版，1993年。
10. 微分流形和黎曼流形。GTM第160号。Springer Verlag，第三版，1995年。
11. Serge Lang.真实和功能分析。GTM 142。Springer Verlag，第三版，1996年。
12. Serge Lang.本科分析。UTM。Springer Verlag，第二版，1997年。
13. 彼得·拉克斯。线性代数及其应用。威利，第二版，2007年。
14. 列别捷夫。特殊功能及其应用。多佛，第一版，1972年。
15. 丹尼尔·莱曼和鲁道夫·布库切。启动“a g’eom’etrie”。PUF，第一版，1988年。
16. 大卫·G·鲁恩伯格。矢量空间法优化。威利，第一版，1997年。
17. David G.Luenberger和Yinyu Ye。线性和非线性规划。Verlag，第四版，2016年。
18. 桑德斯麦克巷和加勒特伯克霍夫。代数。麦克米伦，第一版，1967年。
19. M.-P.马尔利亚文。代数交换。应用程序名称。马森，第一版，1985年。
20. Jerrold E.Marsden和J.R.Hughes，Thomas。弹性的数学基础。多佛，第一版，1994年。
21. 威廉·S·梅西。代数拓扑：简介。GTM第56号。Springer Verlag，第二版，1987年。
22. 威廉·S·梅西。代数拓扑学基础课程。GTM第127号。Springer Verlag，第一版，1991年。
23. Jiri Matousek和Bernd Gartner。理解和使用线性规划。大学文本。Springer Verlag，第一版，2007年。
24. Dimitris N.metaxas。基于物理的变形模型。Kluwer学术出版社，第一版，1997年。
25. 卡尔·D·迈耶。矩阵分析和应用线性代数。暹罗，第一版，2000年。
26. 约翰·W·米尔诺。从可微的观点来看拓扑。弗吉尼亚大学出版社，第二版，1969年。
27. R.mneimn'e和F.testard。介绍“A th’eorie des groupes de lie classiques”。赫尔曼，第一版，1997年。
28. 森田志。微分形式的几何。第201号数学专著的翻译。AMS，第一版，2001年。
29. 詹姆斯·R·芒克雷斯。流形分析。艾迪生·韦斯利，1991年。
30. 詹姆斯·R·芒克雷斯。拓扑结构。普伦蒂斯·霍尔，第二版，2000年。
31. 伊万·尼文、赫伯特·S·扎克曼和休·L·蒙哥马利。数论概论。威利，第五版，1991年。
32. 约瑟夫奥鲁克。《计算几何》，剑桥大学出版社，1998年第二版。
33. Christos H.Papadimitriou和Kenneth Steiglitz。组合优化。算法和复杂性。多佛，第一版，1998年。
34. 贝雷斯福德·帕利特。对称特征值问题。暹罗出版物，第一版，1997年。
35. 丹·皮多。几何，综合课程。多佛，第一版，1988年。
36. M.Penna和R.Patterson。射影几何及其在计算机图形学中的应用。普伦蒂斯·霍尔，第一版，1986年。
37. R.Tyrrell Rockafella.公司凸分析。普林斯顿的数学里程碑。普林斯顿大学出版社，1970年。
38. Eug'ene Rouch'e和Charles de Comberose。特性，例如，EOM。高蒂尔·维拉斯，第七版，1900年。
39. 沃尔特·鲁丁。真实而复杂的分析。麦格劳·希尔，第三版，1987年。
40. 沃尔特·鲁丁。功能分析。麦格劳·希尔，第二版，1991年。
41. 皮埃尔·塞缪尔。射影几何。数学本科课本。斯普林格·维拉格，第一版，1988年。
42. 皮埃尔·塞缪尔。代数数论。多佛，第一版，2008年。
43. 乔瓦尼·桑松。正交函数。多佛，第一版，1991年。
44. Berhnard Scho–lkopf和Alexander J.Smola。用谷粒学习。麻省理工出版社，第一版，2002年。
45. Bernhard Scho–lkopf、John C.Platt、John Shawe Taylor和Alex J.Smola。估计高维分布的支持度。神经计算，13:1443–14712001。
46. Bernhard Scho–lkopf、Alex J.Smola、Robert C.Williamson和Peter L.Bartlett。新的支持向量算法。神经计算，12:1207–12452000。
47. 亚历山大·施里耶弗。线性和整数规划理论。威利，第一版，1999年。
48. 劳伦特·施瓦茨。拓扑G’en’erale et analyse foncinelle.科学收藏团。赫尔曼，1980年。
49. 劳伦特·施瓦茨。分析集成电路和拓扑结构。科学收藏团。赫尔曼，1991年。
50. 劳伦特·施瓦茨。分析二。计算微分方程。科学收藏团。赫尔曼，1992年。
51. 劳伦特·施瓦茨。分析三、计算综合。科学收藏团。赫尔曼，1993年。
52. 劳伦特·施瓦茨。分析应用。科学收藏团。赫尔曼，1993年。
53. H.Seifert和W.Threlfall。拓扑学教科书学术出版社，第一版，1980年。
54. 丹尼斯·瑟尔。矩阵、理论和应用。GTM编号216。Springer Verlag，第二版，2010年。
55. 让·皮埃尔·瑟尔。算术课程。数学研究生课本，第7号。斯普林格，第一版，1973年。
56. 伊戈尔·R·沙法雷维奇。基本代数几何1。Springer Verlag，第二版，1994年。
57. 约翰肖泰勒和内洛克里斯蒂亚尼。模式分析的核心方法。剑桥大学出版社，第一版，2004年。
58. 施建波和马立克。标准化切割和图像分割。模式分析与机器智能交易，22（8）：888-9052000。
59. J.-C.Sidler先生。G’eom’etrie投影。兴趣，第一版，1993年。
60. Ernst Snapper和Troyer Robert J.公制仿射几何。多佛，第一版，1989年。
61. 丹尼尔·斯皮尔曼。光谱图理论。Uwe Naumannn和Olaf Schenk，编辑，组合科学计算。CRC出版社，2012年。
62. 哈罗德·M·斯塔克。数论概论。麻省理工出版社，第一版，1994年。第八次印刷。
63. G.W.斯图尔特。关于奇异值分解的早期历史。《暹罗评论》，35（4）：551–5661993年。
64. J.J.斯托克微分几何。威利经典。Wiley Interscience，第一版，1989年。
65. J.斯托尔菲。定向射影几何。学术出版社，第一版，1991年。
66. Eric J.Stollnitz、Tony D.Derose和David H.Salesin。计算机图形学理论和应用的小波。摩根考夫曼，第一版，1996年。
67. Gilbert Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, first edition, 1986.
68. Gilbert Strang. *Linear Algebra and its Applications*. Saunders HBJ, third edition, 1988.
69. Gilbert Strang. *Linear Algebra and Learning from Data*. Wellesley-Cambridge Press, first edition, 2019.
70. Gilbert Strang and Nguyen Truong. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, second edition, 1997.
71. Patrick Suppes. *Axiomatic Set Theory*. Dover, 1972.
72. Donald E. Taylor. *The Geometry of the Classical Groups*. Sigma Series in Pure Mathematics, Vol. 9. Heldermann Verlag Berlin, 1992.
73. Claude Tisseron. *G´eom´etries affines, projectives, et euclidiennes*. Hermann, first edition, 1994.
74. L.N. Trefethen and D. Bau III. *Numerical Linear Algebra*. SIAM Publications, first edition, 1997.
75. Emanuele Trucco and Alessandro Verri. *Introductory Techniques for 3D Computer Vision*. Prentice-Hall, first edition, 1998.
76. B.L. Van Der Waerden. *Algebra, Vol. 1*. Ungar, seventh edition, 1973.
77. J.H. van Lint and R.M. Wilson. *A Course in Combinatorics*. Cambridge University Press, second edition, 2001.
78. Robert J. Vanderbei. *Linear Programming: Foundations and Extensions*. Springer, fourth edition, 2014.
79. Vladimir N. Vapnik. *Statistical Learning Theory*. Wiley, first edition, 1998.
80. O. Veblen and J. W. Young. *Projective Geometry, Vol. 1*. Ginn, second edition, 1938. [178] O. Veblen and J. W. Young. *Projective Geometry, Vol. 2*. Ginn, first edition, 1946.
81. Lucas Vienne. *Pr´esentation alg´ebrique de la g´eom´etrie classique*. Vuibert, first edition, 1996.
82. Frank Warner. *Foundations of Differentiable Manifolds and Lie Groups*. GTM No. 94. Springer Verlag, first edition, 1983.
83. David S. Watkins. Understanding the QR algorithm. *SIAM Review*, 24(4):447–440, 1982.
84. David S. Watkins. The QR algorithm revisited. *SIAM Review*, 50(1):133–145, 2008.
85. Alan Watt. *3D Computer Graphics*. Addison-Wesley, second edition, 1993.
86. Ernst Witt. Theorie der quadratischen Formen in beliebigen Ko¨rpern. *J. Reine Angew. Math.*, 176:31–44, 1936.
87. Stella X. Yu. *Computational Models of Perceptual Organization*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA 15213, USA, 2003. Dissertation.
88. Stella X. Yu and Jianbo Shi. Grouping with bias. In Thomas G. Dietterich, Sue Becker, and Zoubin Ghahramani, editors, *Neural Information Processing Systems, Vancouver, Canada, 3-8 Dec. 2001*. MIT Press, 2001.
89. Stella X. Yu and Jianbo Shi. Multiclass spectral clustering. In *9th International Conference on Computer Vision, Nice, France, October 13-16*. IEEE, 2003.
90. Oscar Zariski and Pierre Samuel. *Commutative Algebra, Vol I*. GTM No. 28. Springer Verlag, first edition, 1975.

1896 *BIBLIOGRAPHY*

[189] Gunter Ziegler. *Lectures on Polytopes*. GTM No. 152. Springer Verlag, first edition, 1997.