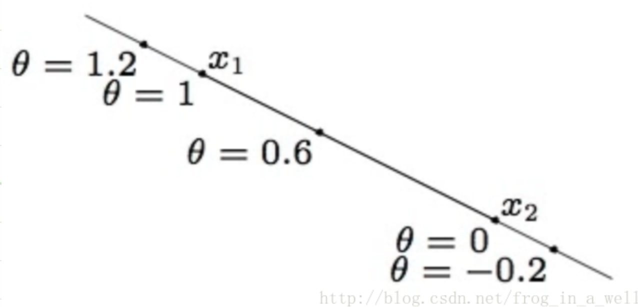
凸优化（convex optimization）第二讲：convex set

⋆ Affine set

affine set表示经过两点的一条线，这条线满足：

$$x = \theta x_1 + (1-\theta)x_2 \qquad (\theta \in \mathbf{R})$$



相较于后面我们要讨论的convex set，这里少了一些限制，是一个更广意义的直线，当θ取任意实数时，这条直线在无限的空间中延伸扩展。

当我们取θ大于1时，此时，x_1占的权重较大，x_2较小，点就在靠近x_1的方向上，当θ小于0 时，此时x_2 点的权重较大，点就在靠近x_2的方向上。

DEF: Affine set标准定义：

包含了所有经过affine set中任意两点的直线的set。

affine set: contains the line through any two distinct points in the set

EX: 什么是affine set:

$$\{x|Ax = b\}$$

方程级的解空间。

如果我们有两个点可以解决上面的方程，那么他们的解的组合也是方程的解：推倒过程如下：

$$A(\theta x_1 + (1-\theta)x_2) = \theta A x_1 + (1-\theta)Ax_2 = \theta b + (1-\theta)b = b$$

根据定义，我想在线性代数中的column space, row space, null space, left null space 也是affine set, 因为他们满足任意两个向量的线性组合也在这个space中。

⋆ Convex set

Convex set就是在affine set的基础之上多了一些条件，那条直线，变成了线段，我们通过限定θ的取值范围来限定set的取值为两点之间的线段。数学描述如下：**line segment** between x_1 and x_2: all points

$$x = \theta x_1 + (1-\theta)x_2$$

with $0 \le \theta \le 1$ **convex set**: contains line segment between any two points in the set

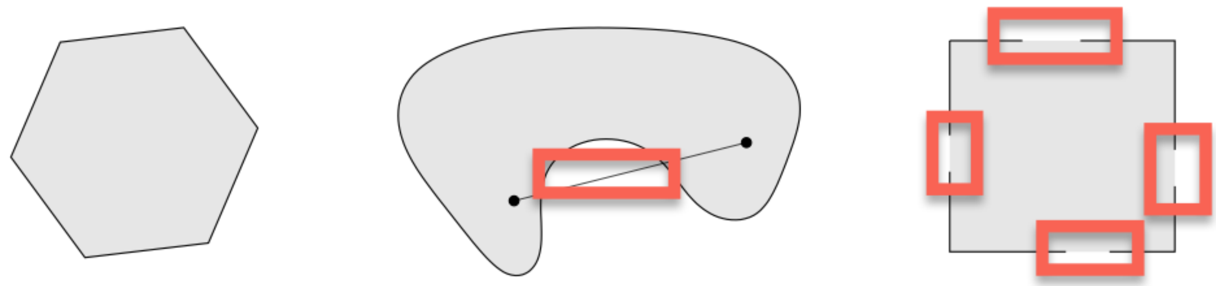
$$x_1, x_2 \in C, 0 \le \theta \le\Rightarrow \theta x_1 + (1-\theta)x_2 \in C$$

通过限定θ取值范围为0到1，我们能够得到：

$$x_1 \le \theta x_1 + (1-\theta)x_2 \le x_2 \quad (x_1 \le x_2)$$

反之亦然。

下面是一些关于convex set的例子：



第一个图是一个正六边形，它是convex的，因为它内部的任意两点的连线也被包含在该图形的内部。

第二图不是convex的，因为当我们连接set中的两点时，出现了未被包含在set的点，即上图中红色方框中的区域。

第三个图不是convex的，原因同二。

⋆ Convex combination

Convex combination是符合以下要求的形式：

convex combination of $x_1, \ldots, x_k$: any point x of the form

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$

with $\theta_1 + \cdots + \theta_k = 1$, $\theta_i \ge 0$

表示的是k个点以不同的权重相加生成一个新的点，权重是非负的，且θ相加之和为1。

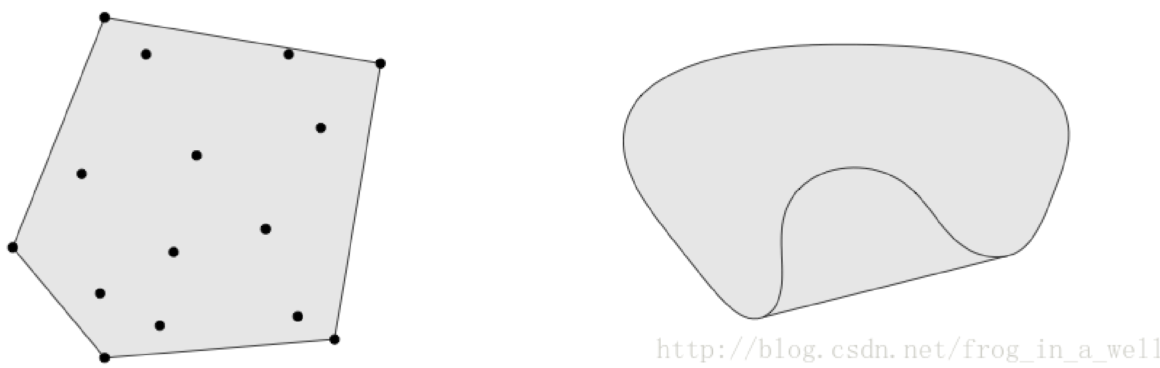
⋆ Convex hull

Convex hull包含了set中所有点的convex combination组合出的所有的新点。其数学表示如下：

$$\textbf{conv}\,C = \{\theta_1 x_1 + \ldots \theta_k x_k \mid x_i \in C, \theta_i \ge 0, i = 1, \ldots, k, \theta_1 + \cdots + \theta_k = 1\}$$

我们再来看看几何直观上的感受：

第一个图表示的是set中的点集通过convex combination过程形成的一个灰色区域，而这个区域就是convex hull。



第二个图是原来的肾形区域(non-convex)通过convex combination过程形成的另外一个灰色区域(convex)。

这个由基础点集生成一个convex空间的概念类似于线性代数中的column vector通过线性的运算来span column space。

那convex hull的作用是什么呢？**就是将non-convex set变成最小的包含原来non-convex set的convex set**。

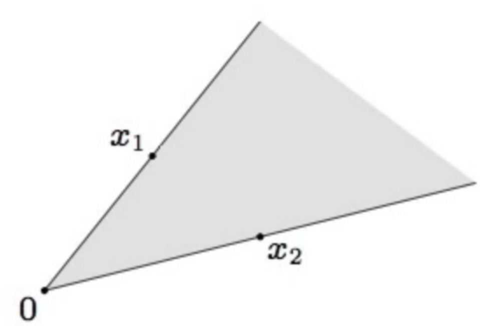
⋆ Conic combination

conic combination指的是由set中的任意两点通过线性组合产生一个新点，线性组合是非负的。

conic (nonnegative) combination of x_1 and x_2: any point of the form

$$x = \theta_1 x_1 + \theta_2 x_2$$

with $\theta_1 \geq 0$, $\theta_2 \geq 0$



几何意义是一个通过vector x_1，x_2产生锥形区域。在线性代数中即为两个vector的线性组合生成的空间中非负组合的那一块。

⋆ Conic cone

Conic cone是包含原来set中所有点经过conic combination 组合产生的点的set。
　　这个set对应至前面我们的affine set-convex set-convex hull 概念链条中的affine set，是一个比较general的概念。

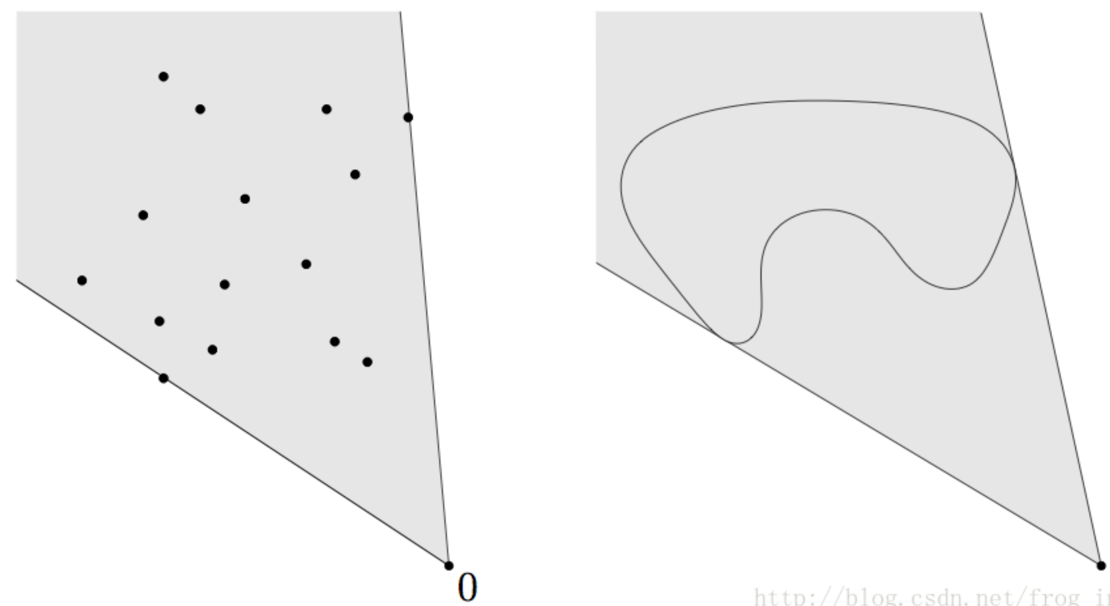
⋆ Conic hull

conic hull是包含了原来set中所有的最小的conic cone。

$$\{\theta_1 x_1 + \cdots + \theta_k x_k | x_i \in C, \theta_i \geq 0, i = 1, \ldots, k\}$$

EX: 例子：

以下是两个conic hull的例子：conic cone可以想象为在以下的hull基础上增加了一些元素。



⋆ 一些重要的例子

通过之前的定义，我们不难得出以下结论：

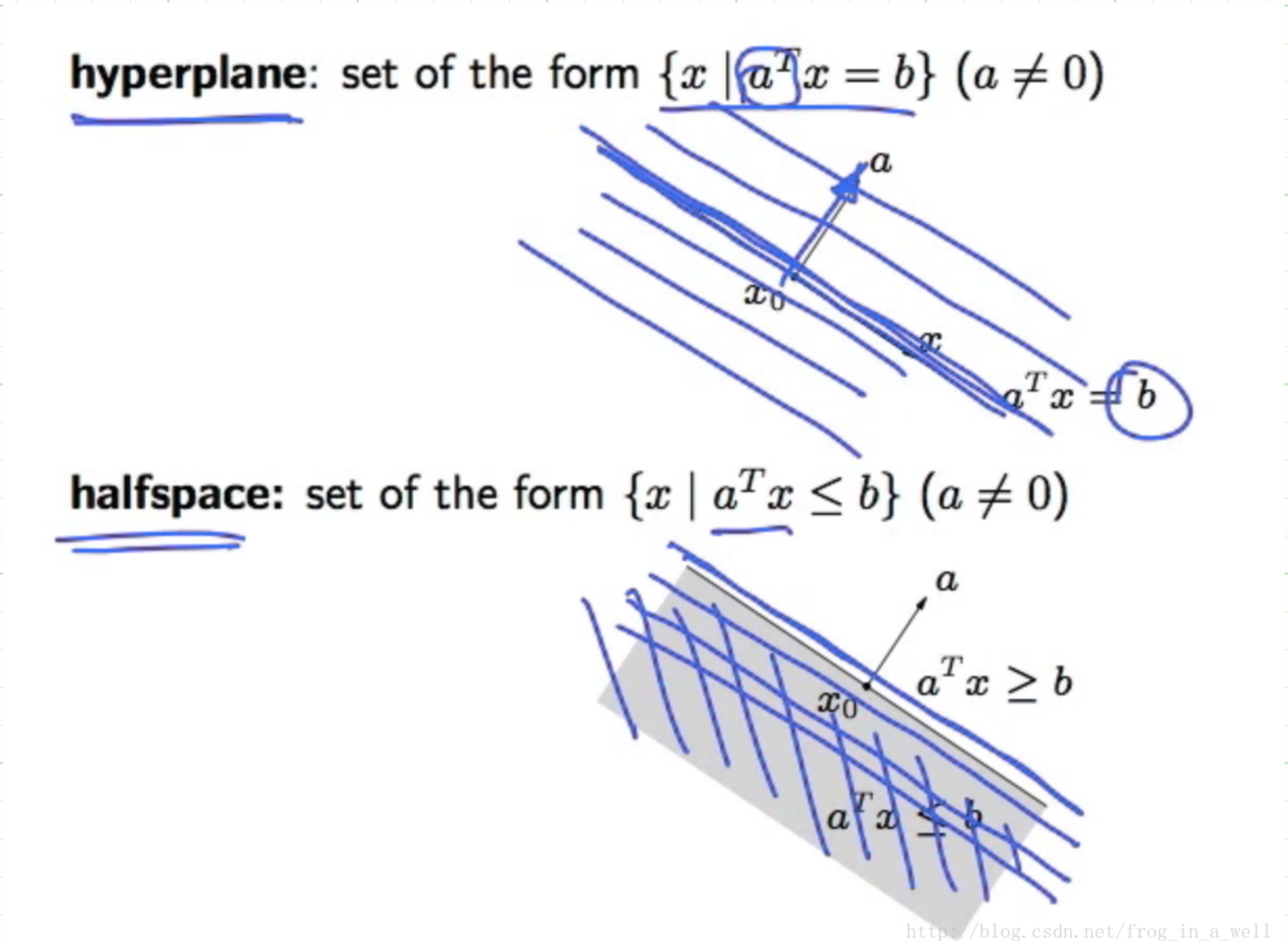
1. 空集和包含一个占的点集是affine且convex的。
2. 一条直线是affine的，如果这条线过原点，那么它还是convex cone。
3. 一条线段是convex的，但不是affince的(除非其为一个点)。
4. 一条射线convex的，但不是affine的。

$$\{x_0 + \theta v | \theta \geq 0\}$$

5. 任何的subspace是affine的，也是convex cone。

⋆ 超平面和半空间

关于超平面，这里有详细的讲解：http://bubblexc.com/y2011/310/



这里要注意的是a和b的值，通过调整b的值我们可以调整超平面距离原点的距离，调整a的值来调整超平面的方向。

超平面是affine和convex的。

半空间是将超平面中的等式改变为不等式，用所确定的超平面来分离空间，使之成为两个半空间。

半空间是convex的，但不是affine的。

⋆ 欧拉球和椭球

DEF: 欧拉球：

表示为离中心点距离小于等于r(半径)的点的集合，数学表示如下：

当然也能表示成另外一种形式：中心点加上一个norm小于1 的向量的r倍。

DEF: 椭球

椭球的公式如下：

P是一个对称正定矩阵，椭球的轴长，由正定矩阵的特征值的平方根给出。

另外一种表示方式是：

unit Ball经过线性变换加上改变中心点形成了一个椭球。

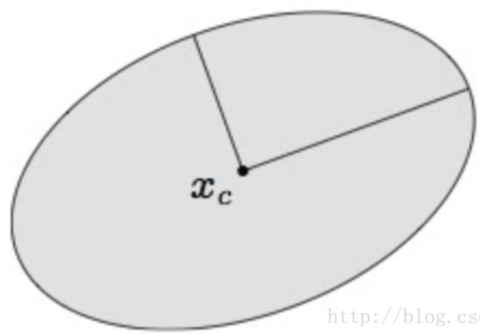
(Euclidean) ball with center x_c and radius r:

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \le r\} = \{x_c + ru \mid \|u\|_2 \le 1\}$$

ellipsoid: set of the form

$$\{x \mid (x - x_c)^T P^{-1}(x - x_c) \le 1\}$$

with $P \in \mathbf{S}^n_{++}$ (*i.e.*, P symmetric positive definite)



other representation: $\{x_c + Au \mid \|u\|_2 \le 1\}$ with A square and nonsingular

Norm balls and norm cones

norm: a function $\|\cdot\|$ that satisfies f( )

- $\|x\| \ge 0$; $\|x\| = 0$ if and only if $x = 0$
- $\|tx\| = |t| \, \|x\|$ for $t \in \mathbf{R}$
- $\|x + y\| \le \|x\| + \|y\|$

notation: $\|\cdot\|$ is general (unspecified) norm; $\|\cdot\|_{\text{symb}}$ is particular norm

norm ball with center x_c and radius r: $\{x \mid \|x - x_c\| \le r\}$

norm是一种function的表示，这种function有以下性质：

1. x的norm大于等于0，当且仅当$x = 0$时，x的norm等于0。
2. 一个向量x的t倍的norm等于t倍的x的norm

3. 两个向量的norm符合三角不等式定理，其几何意义如有上图的平行四边形法则。

Norm外面的symb表示的是symb次方的symb次根。

⋆ Norm Ball和norm cone

Norm Ball表示的是离center点一定距离r的点。

Norm cone表示的是norm的值小于t的点，对于二维的平面，随着t的增大，圆形是逐渐变大的，见下面的圆锥体即为二维平面的norm cone。

Norm balls and norm cones

norm: a function $\|\cdot\|$ that satisfies f( )

- $\|x\| \geq 0$; $\|x\| = 0$ if and only if $x = 0$
- $\|tx\| = |t| \, \|x\|$ for $t \in \mathbf{R}$
- $\|x + y\| \leq \|x\| + \|y\|$

notation: $\|\cdot\|$ is general (unspecified) norm; $\|\cdot\|_{\text{symb}}$ is particular norm

norm cone是convex的。

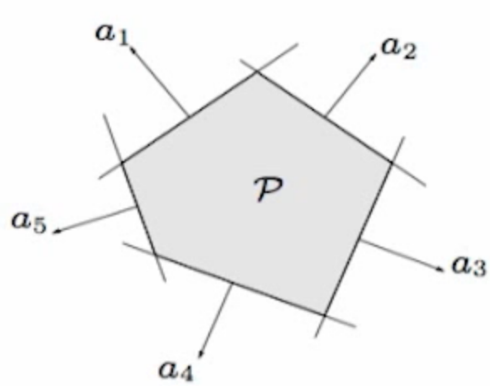
⋆ Polyhedra多面体

Polyhedral在几何上的意义是表示有限个线性不等式或等式的解的可行域，数学表示如下：

solution set of finitely many linear inequalities and equalities

$$Ax \preceq b, \qquad Cx = d$$

($A \in \mathbf{R}^{m \times n}$, $C \in \mathbf{R}^{p \times n}$, $\preceq$ is componentwise inequality)



要注意的是A是$m*n$个元素的矩阵，表示的是m个不等式，每个不等式中有n个元素(constrain variable)。C 表示p个含有n个限制变量的等式。以及挨个元素对比大小的符号，是表示比较两个vector中的每个元素的大小。

其中Ax中的每一行表示的是一个超平面，每个不等式表示的是一个半空间，而整个可行的区域是由以上的等式和不等式共同决定的，就是这些半空间（不等式）和超平面（等式）交集。这些都是线性规划的内容。

⋆ Positive semidefinite cone

notation:

- $\mathbf{S}^n$ is set of symmetric $n \times n$ matrices
- $\mathbf{S}^n_+ = \{X \in \mathbf{S}^n \mid X \succeq 0\}$: positive semidefinite $n \times n$ matrices

$$X \in \mathbf{S}^n_+ \iff z^T X z \geq 0 \text{ for all } z$$

$\mathbf{S}^n_+$ is a convex cone

- $\mathbf{S}^n_{++} = \{X \in \mathbf{S}^n \mid X \succ 0\}$: positive definite $n \times n$ matrices

S表示的是对称的矩阵，下面如果有+号，表示的是对称半正定矩阵，如果有两个加号就是对称正定矩阵。

下图示例展示了一个半正定矩阵确定的cone：

关于用矩阵来表示二次多项式，有如下形成：

$$x^T A x \text{ in } \mathbf{R}^2 \qquad ax^2 + 2bxy + cy^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

对应至二次多项式，上图中的条件就好理解了，具体的推导可以参见Gilbert strang的《Linear Algebra and Its Applications 4ed》第344页。

对应到多维的是这样：

$$x^T A x \text{ in } \mathbf{R}^n \qquad \begin{bmatrix} x_1 & x_2 & \cdot & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdot & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{bmatrix} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

⋆ Operations that preserve convexity保持凸性的运算交集

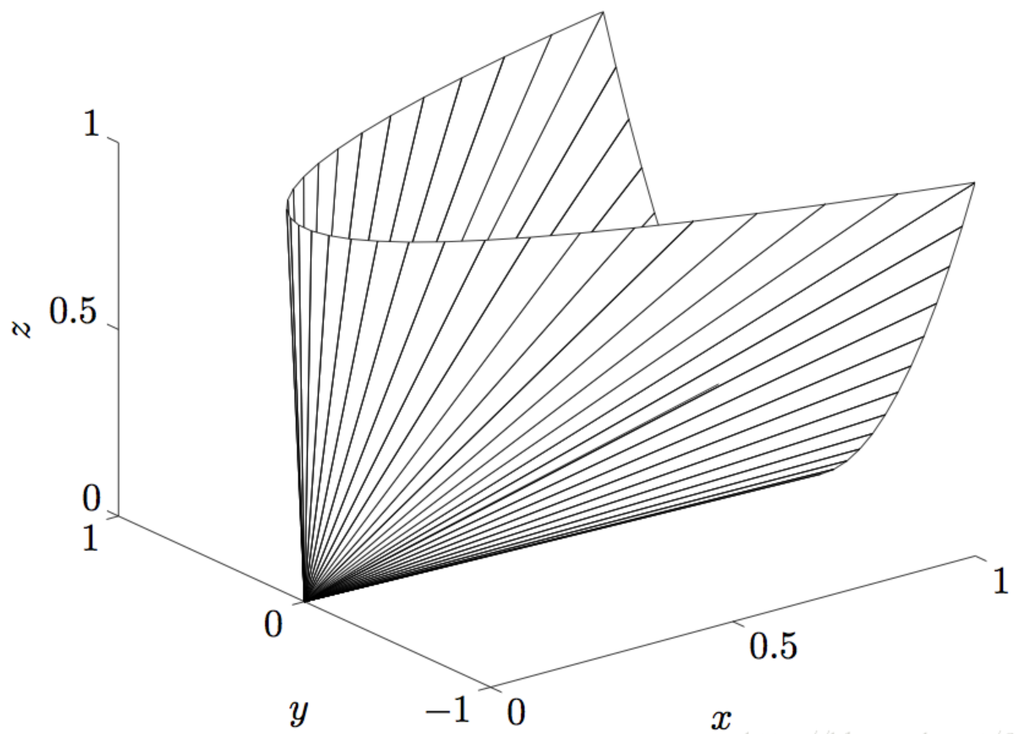
有限个凸集的交集一定是凸集。扩展到无限多的凸集也一样。

EX: 例子：

多边形有限个半空间和超平面的交集，因为半空间和超平面是凸的，那么多边形也是凸的。

所有的凸集S都是一系列半空间的交集。数学表示如下：

$$S = \bigcap \{\mathcal{H} | \mathcal{H} \text{ halfspace}, S \subseteq \mathcal{H}\}$$



EX: 例子：

Example 2.8 We consider the set

$$S = \{x \in \mathbf{R}^m \mid |p(t)| \leq 1 \text{ for } |t| \leq \pi/3\}, \tag{2.10}$$

where $p(t) = \sum_{k=1}^m x_k \cos kt$. The set S can be expressed as the intersection of an infinite number of *slabs*: $S = \bigcap_{|t|\leq\pi/3} S_t$, where

$$S_t = \{x \mid -1 \leq (\cos t, \ldots, \cos mt)^T x \leq 1\},$$

and so is convex. The definition and the set are illustrated in figures 2.13 and 2.14, for $m = 2$.

对于上例，我们可以将$p(t)$看做是两个vector的内积，一个vector是x，另外一个是cos kt，那么这样一个内积确定的超平面的绝对值小于1，即$p(t)$小于等于1，大于等于负1。$p(t)$小于等于1确定了一个半空间，大于等于负1确定了另外一个半空间(halfspace)，最终我们得到的是两个半空间的交集，因为半空间是凸的，所以我们得到的最终结果也是凸的。其几何意义表示如下：

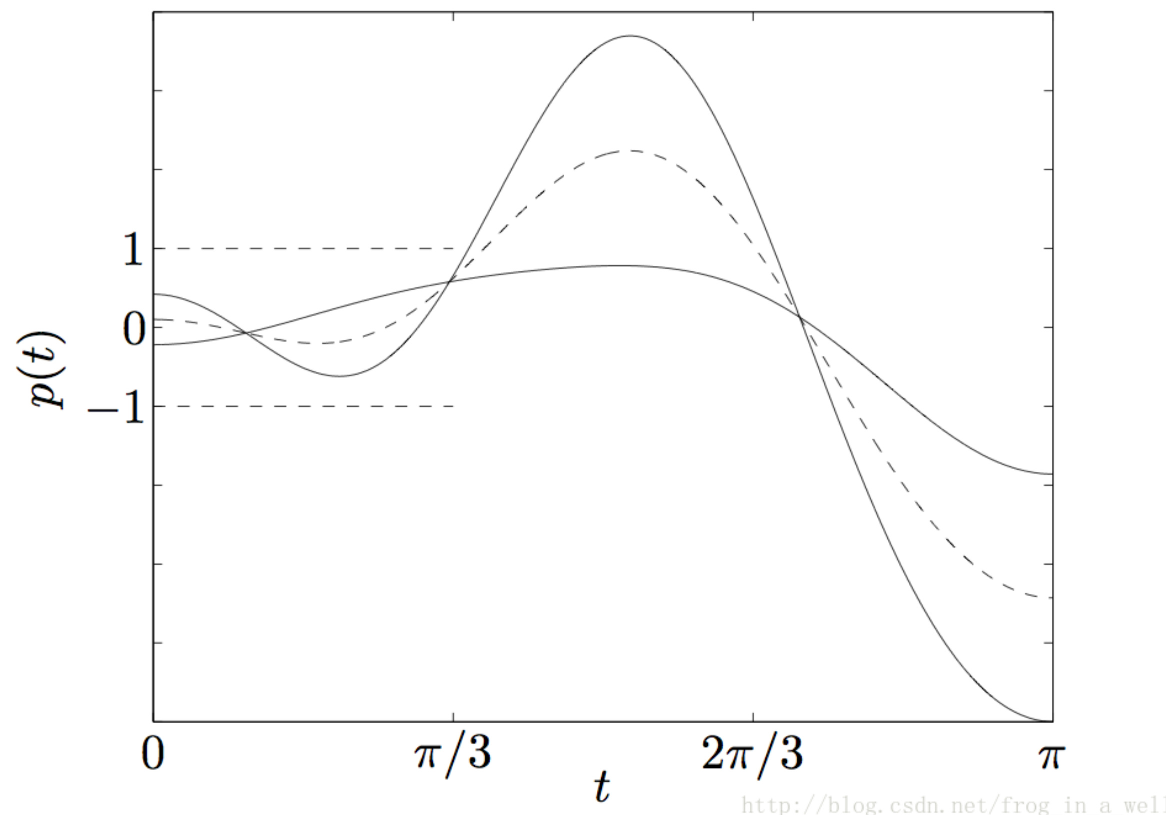
上图可以看出最终的区域是所有满足不等式的vector x所确定的半空间的交集，即为中间区域S，是convex的。

⋆ Affine function 仿射函数

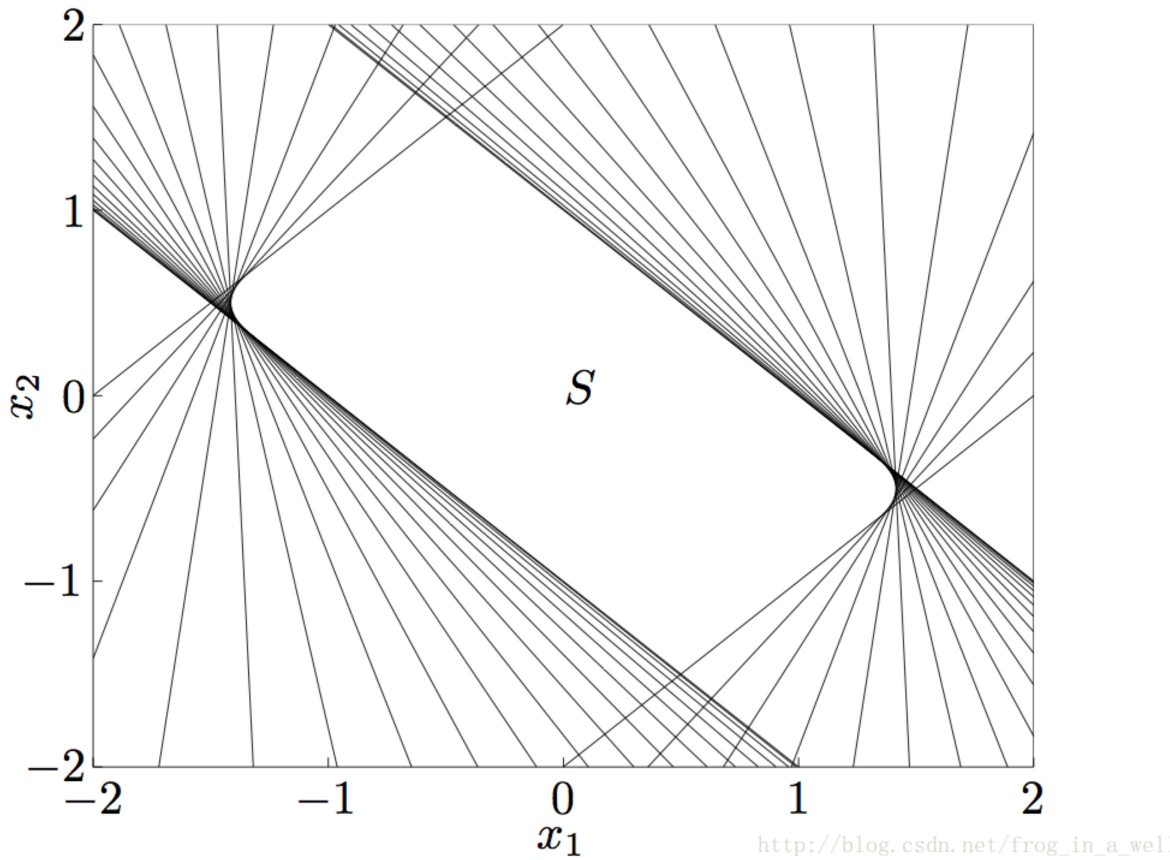
将n维的convex set映射至m维的运算结果也是convex的。

一个通过映射得到的convex set，它的反函数也是convex set。

以下运算也可以保持凸性：



1
0
−1
$p(t)$
0
$\pi/3$
$2\pi/3$
π
t
http://blog.csdn.net/frog_in_a_well



2
1
0
−1
−2
x_2
−2
−1
0
1
2
x_1
S
http://blog.csdn.net/frog_in_a_well

suppose $f: \mathbf{R}^n \to \mathbf{R}^m$ is affine ($f(x) = Ax + b$ with $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$)

- the image of a convex set under f is convex

$$S \subseteq \mathbf{R}^n \text{ convex} \implies f(S) = \{f(x) \mid x \in S\} \text{ convex}$$

$f(S) = \{ f(x) \mid x \in S \}$

- the inverse image $f^{-1}(C)$ of a convex set under f is convex

$$C \subseteq \mathbf{R}^m \text{ convex} \implies f^{-1}(C) = \{x \in \mathbf{R}^n \mid f(x) \in C\} \text{ convex}$$

scaling

$$\alpha S = \{\alpha x | x \in S\}$$

translation

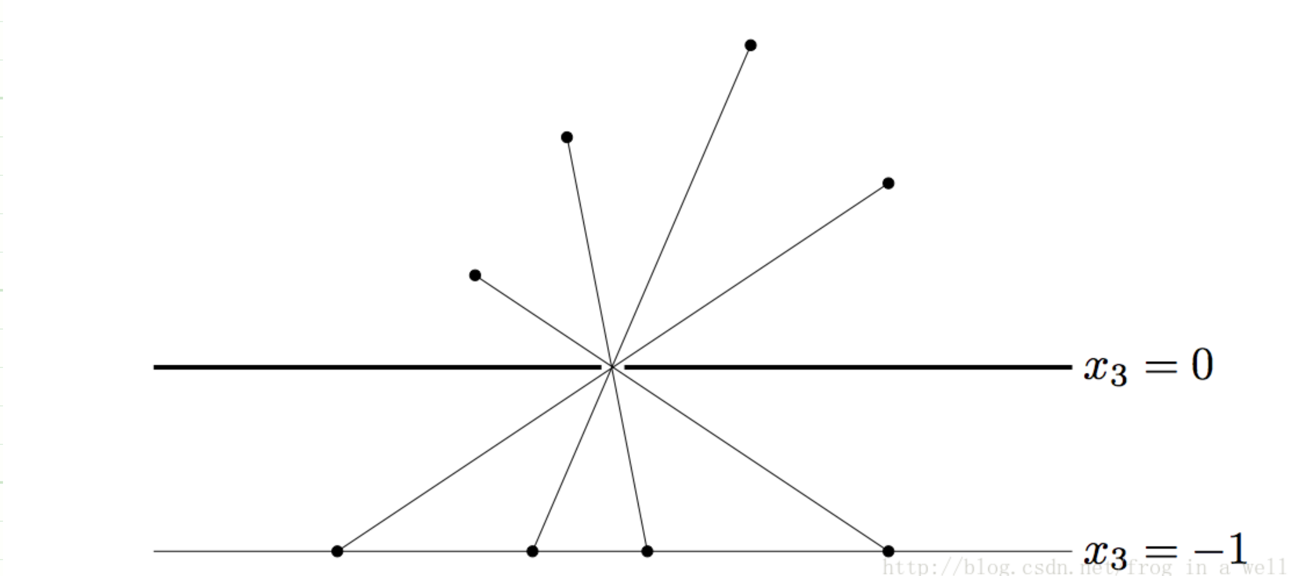
$$S + a = \{x + a | x \in S\}$$

projection

$$T = \{x_1 \in \mathbf{R}^m | (x_1, x_2) \in S \text{ for some } x_2 \in \mathbf{R}^n\}$$

⋆ 透视函数

透视函数的直观理解如下：$x_3 = 0$为一个不透光隔板，只有在原点有一个小孔，光透过小孔投射至$x_3 = -1$的隔板上，如果上面的光点所构成的set是convex的，那么投射至下方的光点构成的set也是convex的。更加直观的想象我们通过一个针孔照相机去拍摄一个convex的对象，那么生成的图像也是convex的。



数学描述如下：

⋆ generalized inequality

proper cones

当K满足如下条件，我们说K是proper cone

我们用这样的proper cone来定义generalize inequality。

If $C \subseteq \mathbf{dom}\, P$ is convex, then its image

$$P(C) = \{P(x) \mid x \in C\}$$

- K is convex.
- K is closed.
- K is *solid*, which means it has nonempty interior.
- K is *pointed*, which means that it contains no line (or equivalently, $x \in K,\ -x \in K \implies x = 0$).

$$x \preceq_K y \iff y - x \in K.$$

y大于等于x，表示成为y与x的差值属于一个proper cone K。
如果y严格大于x，那么表示成为以下：

$$x \prec_K y \iff y - x \in \mathbf{int}\, K,$$

对于矩阵的元素间的比较，有：

- matrix inequality ($K = \mathbf{S}^n_+$)

$$X \preceq_{\mathbf{S}^n_+} Y \iff Y - X \text{ positive semidefinite}$$

⋆ Generalized inequality属性

以上的比较全为vector间的元素比较。

- 加持性：y大于x，v大于u，那么$y+v$大于$x+u$。
- 传递性：y大于x，z大于y，那么z大于x。

- $\preceq_K$ *is preserved under addition*: if $x \preceq_K y$ and $u \preceq_K v$, then $x+u \preceq_K y+v$.
- $\preceq_K$ *is transitive*: if $x \preceq_K y$ and $y \preceq_K z$ then $x \preceq_K z$.
- $\preceq_K$ *is preserved under nonnegative scaling*: if $x \preceq_K y$ and $\alpha \geq 0$ then $\alpha x \preceq_K \alpha y$.
- $\preceq_K$ *is reflexive*: $x \preceq_K x$.
- $\preceq_K$ *is antisymmetric*: if $x \preceq_K y$ and $y \preceq_K x$, then $x = y$.
- $\preceq_K$ *is preserved under limits*: if $x_i \preceq_K y_i$ for $i = 1, 2, \ldots$, $x_i \to x$ and $y_i \to y$ as $i \to \infty$, then $x \preceq_K y$.

- 非负scaling：y大于x，a大于0，那么$a*x$小于$a*y$。
- 反身性：x小于等于x中的元素。
- 反对称性：如果x大于等于y，y大于等于x，那么x等于y。
- 极限保持性：如果x_i小于等于y_i，对于i等于1，2，3直到无穷大都成立，那么x小于等于y。

对于严格小于，有以下性质：

- if $x \prec_K y$ then $x \preceq_K y$.
- if $x \prec_K y$ and $u \preceq_K v$ then $x+u \prec_K y+v$.
- if $x \prec_K y$ and $\alpha > 0$ then $\alpha x \prec_K \alpha y$.
- $x \not\prec_K x$.
- if $x \prec_K y$, then for u and v small enough, $x+u \prec_K y+v$.

⋆ 极小值和最小值：

符号$\preceq$，$\prec$是继承至传统的小于符号，我们已经在上面看到了不传统的大小比较性质可以直接继承用来方便我们的表示，但是有些性质主不能适用了，比如传统的比较是线性的排序，任何的两个元素都是可以比较的，比如两个元素x，y，要不就是x大于等于y，要不就是y大于等于x，总能分出个胜负，但是当我们用于generalized inequality时，大小就不那么明显了。

DEF: Minimum

Minimum的定义如下：

$x \in S$ is **the minimum element** of S with respect to $\preceq_K$ if

$$y \in S \quad \Longrightarrow \quad x \preceq_K y$$

S中的任意一点y都大于等于x，那么x就是S中的minimum点。Minimum我们一般认为是最小值点，即在一个set中是唯一的。但是minimal极小值点可以是多个的。

⋆ Minimal

极小值minimal的定义如下：

$x \in S$ is **a minimal element** of S with respect to $\preceq_K$ if

$$y \in S, \quad y \preceq_K x \quad \Longrightarrow \quad y = x$$

极小值就是在S中找不出来小于或者等于x却又不是x的元素。
用传统的表示方法如下：
对于最小值x，有：

$$S \subseteq x + K$$

这里的$x + K$表示的是所有大于等于x的值，K是proper cone。
对于极小值x，有：

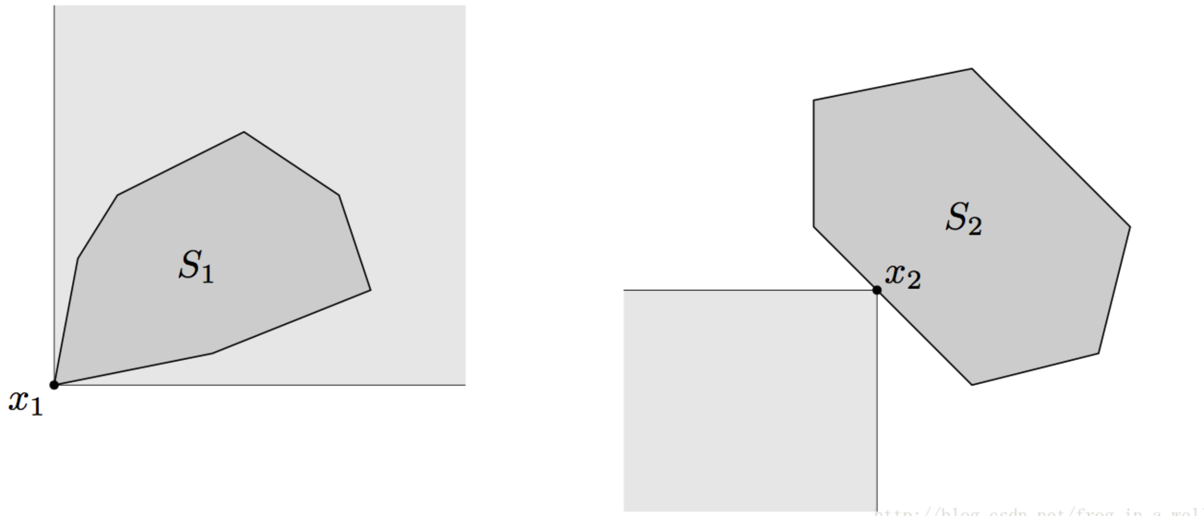
$$(x - K) \cap S = \{x\}$$

表示的是所有小于x的值与S的交集只有x自己。
对于最小值minimum和极小值minimal的直观表示如下：
左边的图深色区域表示的是convex set S_1，浅灰色区域是最小值$+K$，那么我们的任务是从S_1中找出一个点使得S_1中所有的点都大于等于它，这个点就是x_1。右边的图深色区域表示的是convex set S_2，浅灰色区域是$x - K$，即小于x的值，根据定义小于极小值的值与S的交点只有极小值，那么极小值就是x_2。这里我们发现极小值并不唯一，因为当我们取x_2所在直线边缘的其他点，也满足这样的要求。

⋆ Separating hyperplane theorem超平面分割理论

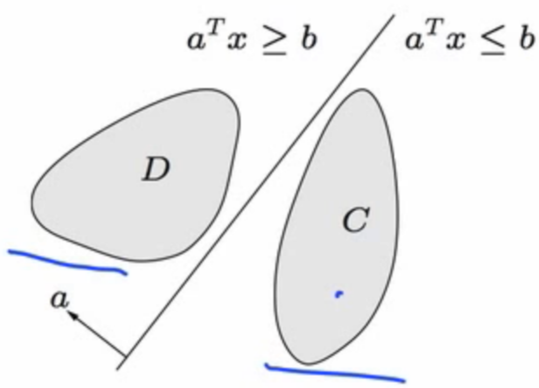
DEF: Separating hyperplane

对于两个在空间中不相交的set，我们总能找到合适的超平面来分割它们，图形如下所示：
严格的分割面只是将上述的大小等于或者小于等于符号换成大于或者小于。



if C and D are disjoint convex sets, then there exists $a \neq 0$, b such that

$$a^T x \leq b \text{ for } x \in C, \qquad a^T x \geq b \text{ for } x \in D$$



the hyperplane $\{x \mid a^T x = b\}$ separates C and D

DEF: Support hyperplanes

支持超平面就是经过set边缘的超平面，对于convex set，在set的边缘的每一点都有这样的支持超平面。

这里要注意的是有些点的支持超平面不唯一，特别是像一些不可导的点，想象下正方形的四个角。

★ dual cone and generalized inequality

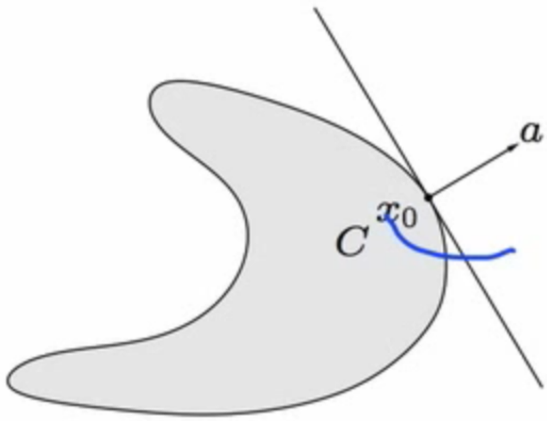
DEF: dual cone

dual cone的定义如下：假设x属于一个cone K，那么dual cone就是那些与所有x相乘大于等于0的向量所构成的cone。

supporting hyperplane to set C at boundary point x_0:

$$\{x \mid a^T x = a^T x_0\}$$

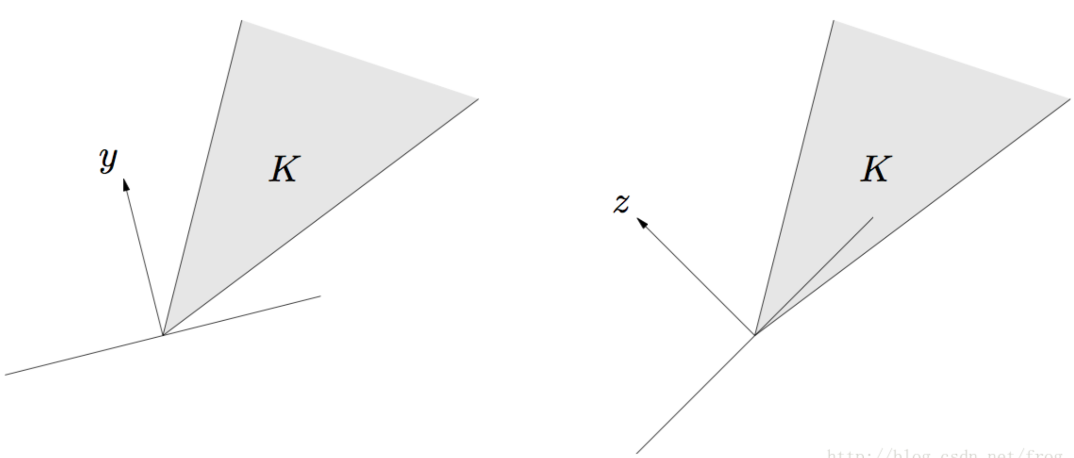
where $a \neq 0$ and $a^T x \leq a^T x_0$ for all $x \in C$



dual cone of a cone K:

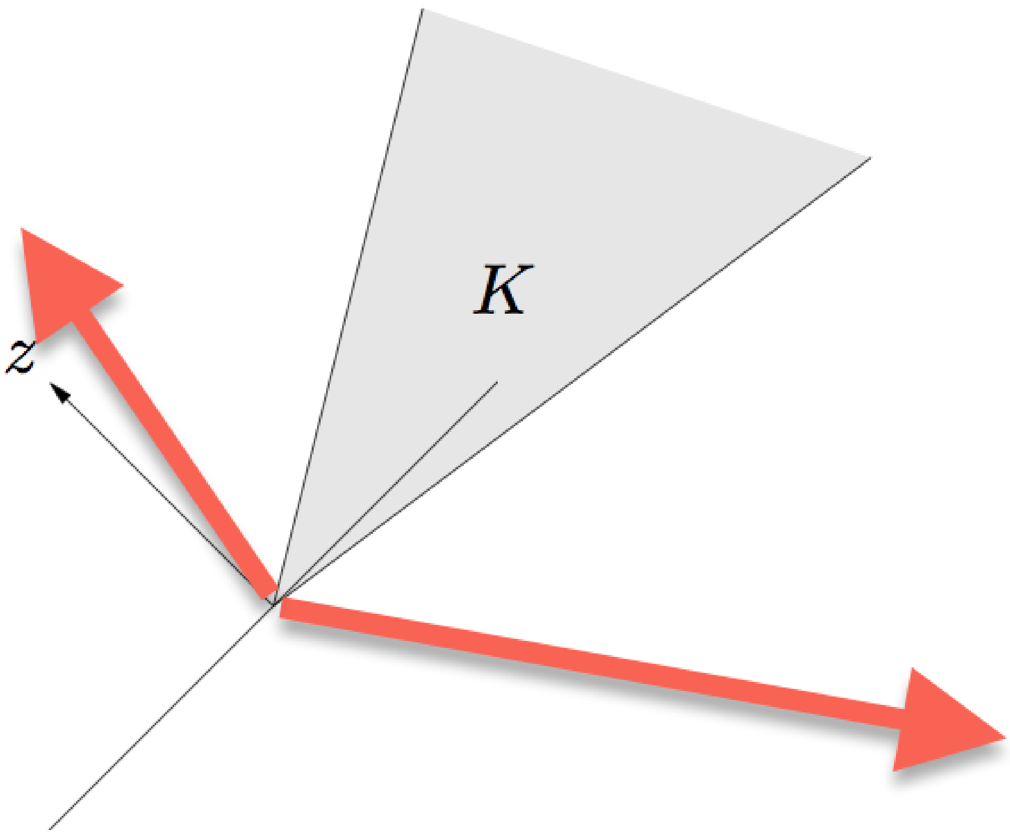
$$K^* = \{y \mid y^T x \geq 0 \text{ for all } x \in K\}$$

直观的感受如下：



左图的y与K中的所有向量的夹角小于90度，所以根据乘积法则，它们的乘积始终大于0，所以y属于K的dual cone中的一员。右图的z与K中的部分向量的夹角大于90 度，所

以部分的乘积小于0，则z不属于K的dual cone。So，我们画出它的dual cone是如下样子滴，两条红线所夹的区域两条红色的线是分别与K的边界成90度的线，如果大于了90度就会出现与原来cone乘积小于0的情况。



所以我们可以想象一下，如果original cone K是比较苗条的，那么dual cone就比较粗胖。如果original cone是二维空间的一个half space，那么dual cone就是一条与分割超平面垂直的向量。

下面是一些例子：

examples

- $K = \mathbf{R}^n_+$: $K^* = \mathbf{R}^n_+$
- $K = \mathbf{S}^n_+$: $K^* = \mathbf{S}^n_+$
- $K = \{(x,t) \mid \|x\|_2 \le t\}$: $K^* = \{(x,t) \mid \|x\|_2 \le t\}$
- $K = \{(x,t) \mid \|x\|_1 \le t\}$: $K^* = \{(x,t) \mid \|x\|_\infty \le t\}$

如果K是第一象限，那么其dual cone也是第一象限，推广至任何成90度的cone的dual cone都是它们自身，这也叫做自身对偶性。

半正定矩阵的dual cone也是半正定矩阵。