5. 凸优化问题的定义

朱天宇

1 凸的定义

1.1 仿射集 Affine set

如果一个集合的任意两个元素的线性组合都在这个集合内，称这个集合为仿射集[1]，即对于集合 C，有：

$$\forall x, y \in C, \theta \in \mathbb{R}, \rightarrow \theta x + (1-\theta)y \in C$$

这个定义很容易推广到多个元素的情况，即

$$\forall x_1, x_2, ..., x_n \in C, \theta_1, \theta_2, .., \theta_n \in \mathbb{R}, \rightarrow \theta_1 x_1 + \theta_2 x_2 + ... + \theta_n x_n \in C$$

其中 $\theta_1 + ... + \theta_n = 1$。$\theta_1 x_1 + \theta_2 x_2 + ... + \theta_n x_n$ 被称为仿射组合

1.2 子空间

对于一个仿射集 C 以及其内部任意一点 x_0，定义集合 V

$$V = C - x_0 = \{x - x_0 | x \in C\}$$

这里 V 是一个子空间。(满足性质：零向量属于子空间，子空间任意两个向量的加和，单向量与标量的乘积也在子空间里)，仿射空间是子空间的一个 平移 。

对于矩阵 $\mathsf{A} \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^n$，其解集 $C = \{x | \mathsf{A}x = b\}$ 是一个仿射集。令集合 $V = \{v | \mathsf{A}v = 0\}$ 是一个子空间，并且有 $\mathsf{A}x_0 = b$，那么就有 $C = V + x_0$。

举例 考虑方程组 $x_1 + 2x_2 = 3$，其含有解 $x_1 = -1, x_2 = 2$ 以及 $x_1 = 1, x_2 = 1$，那么这两个解的线性组合也是原方程组的解。比如 $x_1 = 0.8 * -1 + 0.2 * 1 = -0.6, x_2 = 0.8 * 2 + 0.2 * 1 = 1.8$ 也是方程的解。其齐次形式 $x_1 + 2x_2 = 0$ 的任意解，比如 $x_1 = -2, x_2 = 1$ 加上一个特解 $x_1 = 1, x_2 = 1$，也是原方程的解。

1.3 仿射包 affine hull

一个集合 C 内所有点的仿射组合，被称为仿射包

$$\mathsf{aff}C = \{\theta_1 x_1 + ... + \theta_k x_k | x_1, ..., x_k \in C, \theta_1 + ... + \theta_k = 1\}$$

仿射包 aff$C$ 是包含集合 C 中所有元素的最小的仿射集。

1.4 凸集 convex set

对于集合 C，如果有：

$$\forall x, y \in C, \theta \in \mathbb{R}, \theta > 0 \rightarrow \theta x + (1-\theta)y \in C$$

对比仿射集的定义，凸集要求系数 $\theta > 0$。同样可以推广到任意个点的情况。同样，仿射组合也对应 “凸组合”

1.5 凸包 convex hull

类似的定义，集合 C 的凸包，记为 conv$C$，是包含集合 C 中所有元素的凸组合的集合，是包含集合 C 的所有元素的最小的凸集：

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

[1] https://blog.csdn.net/robert_chen1988/article/details/80048245

1.6 锥 cone

一个集合 C 被称为锥，如果对于 $\forall x \in C, \theta \geq 0 \rightarrow \theta x \in C$

一个集合 C 被称为 凸锥 ，如果对于 $\forall x_1, x_2 \in C, \theta_1, \theta_2 \geq 0, \rightarrow \theta_1 x_1 + \theta_2 x_2 \in C$

锥组合：$\theta_1 x_1 + ... + \theta_k x_k, \theta_1, ..., \theta_k \geq 0$。

锥包 (conic hull): $\{\theta_1 x_1 + ... + \theta_k x_k | x_i \in C, \theta_i \geq 0\}$

下图分别展示了一组离散点的锥包和一个非凸形状的锥包。

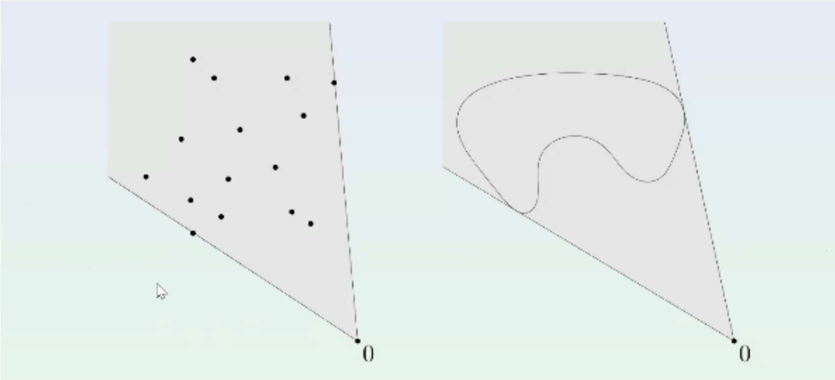

Figure: **Left.** The shaded set is the conic hull of a set of fifteen points (not including the origin). **Right.** The shaded set is the conic hull of the non-convex kidney-shaped set that is surrounded by a curve.

值得注意的是两个图中都绘制了一个原点。

1.7 一些例子和应用

超平面 一个超平面是满足 $\{\mathsf{x} | a^\top \mathsf{x} = b\}$ 的集合，是一个凸集。

半空间 半空间是满足 $\{\mathsf{x} | a^\top \mathsf{x} \leq b\}$ 的集合，也是一个凸集。

多面体 Polyhedra 一个多面体被定义为一组有限的线性等式与不等式的解。当然是凸的。

$$\mathcal{P} = \{\mathsf{x} | a_j^\top \mathsf{x} \leq b_j, j = 1, ..., m, c_k^\top \mathsf{x} = d_k, k = 1, ..., p\}$$

球 Ball $B(\mathsf{x}_c, r) = \{\mathsf{x} | \|\mathsf{x} - \mathsf{x}_c\|_2 \leq r\}$ 这里半径大于 0，范数也可以不是 2。但不可以小于 1，小于 1 无法满足范数的定义所满足的性质 (三角不等式约束)。

范式球与范式锥 [2]

球心和半径都确定的范式球的定义为 $\{\mathsf{x} | \|\mathsf{x} - \mathsf{x}_c\| \leq r\}$

同样，定义范式锥：$C = \{(\mathsf{x}, t) | \|\mathsf{x}\| \leq t\}$

值得注意的是，范式锥比范式球高一维——范式锥的 r 也是维度之一。

如果在矩阵空间中定义，则有一些其他的凸结构例子：

半正定锥 [3] 对称矩阵 S^n, 半正定对称矩阵 S_+^n，正定对称矩阵 S_{++}^n 都是凸的。（详见参考链接）

1.8 正常锥 proper cone 以及广义不等性

定义 正常锥可以用于对锥中的元素定义一个偏序的关系。一个正常锥 K 是满足如下几个条件的锥：

1. 凸的
2. 闭的，即包含边界

[2] https://blog.csdn.net/robert_chen1988/article/details/80479813

[3] https://www.bilibili.com/read/cv4665074

3. 非空的 (solid)，即具有非空内部。
4. 定向的 (pointed)，含义是，锥的内部不能含有直线，例如：$x \in K and -x \in K \Rightarrow x = 0$ 也就是说除非 x 就是原点，否则 x 与 $-x$ 不能同时在集合内部。

根据这个定义，就可以定义一个集合内部的偏序关系。在 $\mathbb{R}^n$ 中的偏序拥有在 $\mathbb{R}$ 中的有序的许多性质。

$$x \preceq_K y \Leftrightarrow y - x \in K$$

$$x \prec_K y \Leftrightarrow y - x \in \text{int}K$$

在 $K = \mathbb{R}_+$ 的时候 (也就是正的数轴)，$\preceq_K$ 和 $\leq$ 是相同含义；

举例

1. n 维的非负象限 (orthant)$K = \mathbb{R}^n_+$ 是一个正常锥。此时 x $\preceq_K$ y 的含义就是 x 的每一个分量都小于 y，即 $x_i < y_i, i = 1...n$。
2. 在由所有的 n 维对称矩阵构成的空间 S^n 中，半正定锥 S^n_+ 是一个正常锥。此时 X $\preceq_K$ Y 的含义是 X − Y 是半正定的。由于经常使用，对于半正定矩阵，可以忽略广义不等 $\preceq_K$ 的下标。X $\preceq$ Y 就可以直接表示一种偏序关系。
3. 在 $[0, 1]$ 区间上的非负多项式，也就是

$$K = \{\mathsf{c} \in \mathbb{R}^n | c_1 + c_2 t + c_3 t^2 + ... + c_n t^{n-1}, t \in [0, 1]\}$$

此时对于两个向量 c, d, c $\preceq$ d 当且仅当 $c_1 + c_2 t + ... + c_n t^{n-1} \leq d_1 + d_2 t + ... + d_n t^{n-1}$

1.9 极小元和最小元

广义不等 $\preceq$ 是不等 $\leq$ 的类比。有些属性是可以类比的，但有些属性不能类比。最常见的就是，在 $\mathbb{R}$ 中，任意两个元素都是 可比 的。因此，在广义不等中，极小和最小的定义就变得复杂。

最小元 minimum 称 x 是 S 中的最小元，如果 $\forall$y $\in S \Rightarrow$ x $\preceq_K$ y，显然，最小元是唯一的。使用**集合符号**来描述的话，就是：x $\in S$ 是最小元当且仅当 $S \subseteq$ x $+ K$。

极小元 minimal 称 x 是 S 中的极小元，如果 y $\preceq$ y $\Rightarrow$ y = x。也就是说，如果 x, y 可比，那么一定有 x $\preceq$ y。用集合符号描述就是：x $\in S$ 是极小元当且仅当 (x $- K) \cap S = \{$x$\}$。

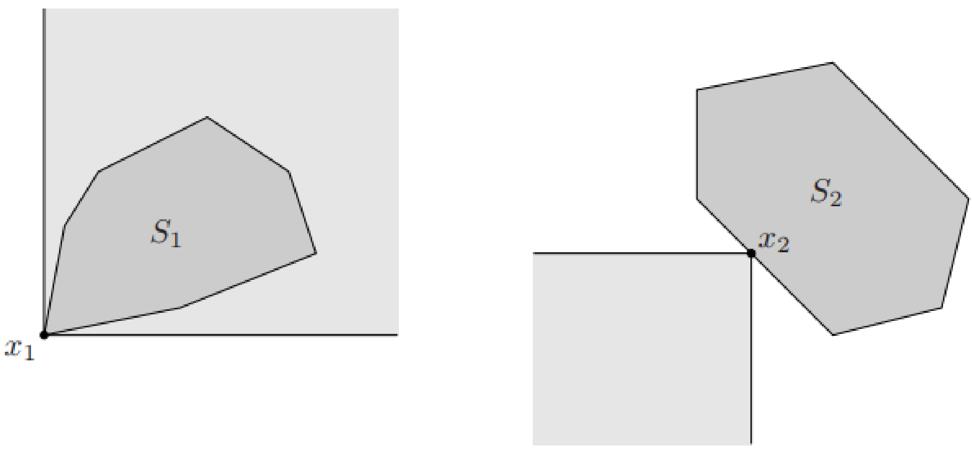

Figure 2.17 *Left.* The set S_1 has a minimum element x_1 with respect to componentwise inequality in $\mathbf{R}^2$. The set $x_1 + K$ is shaded lightly; x_1 is the minimum element of S_1 since $S_1 \subseteq x_1 + K$. *Right.* The point x_2 is a minimal point of S_2. The set $x_2 - K$ is shown lightly shaded. The point x_2 is minimal because $x_2 - K$ and S_2 intersect only at x_2.

2 凸函数 convex funtion

2.1 定义

函数是一个 $\mathbb{R}^n \to \mathbb{R}$ 的一个映射。如果这个映射是凸的，当且仅当

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

也就是凸组合的函数值小于等于函数值的凸组合。严格凸就是当 $x \neq y$ 的时候，$\leq \Rightarrow <$

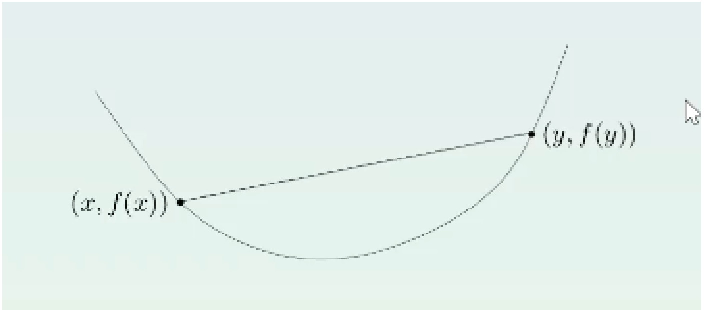

Figure: Graph of a convex function.

梯度 (一阶导) 定义 充要条件：$f(y) \geq f(x) + \nabla f(x)^\top (y-x)$，表示沿着任意一点做切线，函数图像都在切线的上方。

二阶导定义 充要条件：$\nabla^2 f(x) \succeq 0$，也就是 Hessian 矩阵是半正定的。但值得注意的是，Hessian 正定可以推出严格凸，但反之，严格凸无法推导出 Hessian 正定。比如 $f(x) = x^4$，其 Hessian 在 $x = 0$ 处等于 0。（但也满足半正定的要求）。

和前述最优性条件的区分：最优性条件是在特定的点处的性质，且满足二阶最优性条件的前提是满足一阶最优性条件；而凸性的一阶二阶条件是在整个函数上的性质。

2.2 常见的凸函数

1. 指数函数 e^{ax}，在 $\mathbb{R}$ 上。
2. 幂函数 x^a，在 $\mathbb{R}_{++}$ 上，也就是要求 x 是正数；当 $a \geq 1$ or $a \leq 0$ 的时候是凸的，在 $[0,1]$ 上是凹的。
3. 绝对值的幂函数 $|x|^p, p \geq 1$ 在 $\mathbb{R}$ 上是凸的
4. 对数函数在 $\mathbb{R}_{++}$ 上是凸的
5. 熵函数也是凸的 $x\log(x)$，但要注意定义域。(约定 $0\log(0) = 0$)
6. 范数是凸的。
7. 最大值函数 $f(\mathrm{x}) = \max(x_1, x_2, ..., x_n)$ 是凸函数。这个函数无法求导，但是可以从定义角度来证明。
8. 指数求和函数 $f(\mathrm{x}) = \log(e^{x_1} + e^{x_2} + ... + e^{x_n})$ 是凸的。这个函数满足

$$\max\{x_1, ..., x_n\} \leq f(\mathrm{x}) \leq \max\{x_1, ..., x_n\} + \log(n)$$

9. 几何平均 $f(\mathrm{x}) = (\prod_{i=1}^{n})^{1/n}$ 是 凹 的。
10. 行列式的对数函数 $f(\mathrm{X}) = \log\det(\mathrm{x})$ 是凸的。

2.3 常见的凸函数算子

以下介绍一些常见的可以使得函数保持凸性的函数算子。

1. 非负的加权和 $f = w_1 f_1 + ... + w_m f_m, w_1, ..., w_m \geq 0$

2. 推广到无限项相加的情况，即为积分：如果对于 $\forall y \in \mathcal{A}$，都有 $f(x,y)$ 是凸的，且 $w(y) \geq 0$, 那么函数

$$g(x) = \int_{\mathcal{A}} w(y) f(x,y) \mathrm{d}y$$

关于 x 也是凸的。

3. 和仿射变换组合：$f : \mathbb{R}^n \mapsto \mathbb{R}, A \in \mathbb{R}^{n\times m}, b \in \mathbb{R}^n, g : \mathbb{R}^m \mapsto \mathbb{R}$，那么有

$$g(x) = f(\mathrm{A}x + b)$$

如果 f 是凸的，那么 g 也是凸的。

4. 分片最大值：$f(x) = \max\{f_1(x), f_2(x)\}$，如果 f_1, f_2 都是凸的，那么 f 也是凸的。
5. 分片上确界：如果对于 $\forall y \in \mathcal{A}$，都有 $f(x,y)$ 是凸的，那么有

$$g(x) = \sup_{y \in \mathcal{A}} f(x,y)$$

也是凸的。(当然要满足定义域存在、上确界存在等等条件。)

6. Quasi-convex 拟凸函数 $S_\alpha = \{x \in \mathrm{dom} f | f(x) \leq \alpha\}$ 也是凸的 (水平集)。拟凸函数：如果一个函数本身不是凸的，但是其水平集都是凸的，则称其为拟凸函数。[4]
7. 对数凹函数：如果函数 f 满足 $f(x) > 0, \forall x \in \mathrm{dom} f$，且 $\log f$ 是凹的，那么这就是一个对数凹函数。也就是说对于函数本身不要求是凹的，但是其对数是凹的，那么这种类型就是对数凹函数。

[4] https://zhuanlan.zhihu.com/p/131604034