最优化理论

陈鸿峥

2019.03*

目录

1 简介

1.1 优化概述

优化(optimization)：从一个可行解的集合中寻找出最好的元素

- 优化变量$\mathbf{x} \in \mathbb{R}^n$
- 目标/损失函数$f_0 : \mathbb{R}^n \mapsto \mathbb{R}$
- 不等式约束函数$f_i : \mathbb{R}^n \mapsto \mathbb{R}$
- 等式约束函数$h_j : \mathbb{R}^n \mapsto \mathbb{R}$
- 可行解$\mathcal{S} = \{\mathbf{z} \mid f_i(\mathbf{z}) \leq 0, h_j(\mathbf{z}) = 0, i = 1, \ldots, m, j = 1, \ldots, p\}$
- 最优解$\mathbf{x}^\star \iff \forall \mathbf{z} \in \mathbb{R}^n, \mathbf{z} \in \mathcal{S} :\ f_0(\mathbf{z}) \geq f_0(\mathbf{x}^\star)$

$$
\begin{array}{lll}
\text{minimize} & f_0(\mathbf{x}) & \\
\text{subject to} & f_i(\mathbf{x}) \leq 0 & i = 1, \ldots, m \\
& h_j(\mathbf{x}) = 0 & j = 1, \ldots, p
\end{array}
$$

*Build 20190305

例 1.

- 最小二乘线性拟合（凸问题）
- 深度神经网络（非凸，见下）

$$\begin{aligned}\mathbf{x}_1^{(i)} &= f_1(\mathbf{x}_0^{(i)}, \mathbf{w}_1)\\ \cdots &= \cdots\\ \mathbf{x}_n^{(i)} &= f_n(\mathbf{x}_{n-1}^{(i)}, \mathbf{w}_n)\\ \min \sum_{i=1}^{m} &(\mathbf{y}^{(i)} - \mathbf{x}_n^{(i)})^2\end{aligned}$$

- 图像处理，自然图像通常都是**分块光滑**的，原图Φ_0，有噪声的新图Φ
  全变参*(TV, Total Variation)*范数，计算图像每个像素点左侧和下侧的差异

$$\|\Phi\|_{TV} = \sum_y \sum_x \sqrt{(\Phi(x,y) - \Phi(x,y-1))^2 + (\Phi(x,y) - \Phi(x-1,y))^2}$$

  可得优化目标：近似自然图像，而且跟原图不能差太远

$$\min(\|\Phi\|_{TV} + \lambda\|\Phi - \Phi_0\|_F^2)$$

- 推荐系统：*Netflix*问题
  矩阵横向为用户，纵向为电影，值为评分值$(1 \sim 5)$，问题是把矩阵补全，这样就可以做推荐了$\to$低秩矩阵补全
  电影很多，但类型不多，关联关系有限$\to$**近似低秩**[1]
  低秩本来需要最小化$\mathbf{z}$的非零奇异值数目$\|\mathbf{z}\|_0$，但是非凸的；转化为最小化和范数[2]$\|\mathbf{z}\|_\star$

$$\begin{aligned}\min \quad & \|\mathbf{z}\|_\star := \|\mathbf{z}\|_1\\ s.t. \quad & \mathbf{z}_{ij} = \mathbf{M}_{ij},\ (i,j) \in \Omega\end{aligned}$$

1.2 分类

- 线性规划/非线性规划
- 凸规划/非凸规划（更好的分类）

目标函数凸函数，可行解集为凸集则是凸优化，一般容易求解

1.3 历史

- Newton-Raphson算法：求零点，等价于求$\min f^2(x)$
- Gauss-Seidel算法：求解线性方程组$A\mathbf{x} = \mathbf{b}$，等价于求$\min \|A\mathbf{x} - \mathbf{b}\|_2^2$
- Lagrange

[1] A的秩等于非零奇异值$\sqrt{eig(A^\mathrm{T}A)}$数目
[2] 矩阵所有奇异值之和

- Kantoronc：苏联，线性规划，诺贝尔经济学奖
- Dantzig：美国，优化决策，线性规划单纯形
- Von Neumann：线性规划问题对偶理论
- Karmarkar：80年代，线性规划内点法
- Nesterov：后80年代，非线性凸优化内点法
- 现代：并行、随机算法

2 凸集

定义 1. 一些集合概念如下

- 仿射集*(affine set)*

$$
\begin{aligned}
\mathcal{C}\text{为仿射集} &\iff \text{过}\mathcal{C}\text{内任意两点的}\textbf{直线}\text{都在}\mathcal{C}\text{内} \\
&\iff \forall x_1, x_2 \in \mathcal{C}, \theta \in \mathbb{R}, \theta x_1 + (1-\theta)x_2 \in \mathcal{C}
\end{aligned}
$$

  例 2. 用定义易证线性方程组的解集$\mathcal{C} = \{\mathbf{x} \mid A\mathbf{x} = \mathbf{b}\}$是仿射集；反过来，每一个仿射集都可以用线性方程组的解集表示

- 仿射组合

$$
\forall x_1, x_2, \ldots, x_k \in \mathcal{C}, \theta_1, \ldots, \theta_k \in \mathbb{R}, \theta_1 + \cdots + \theta_k = 1 : \ \theta_1 x_1 + \cdots + \theta_k x_k \in \mathcal{C}
$$

- 仿射包*(hull)*：所有仿射组合的集合

$$
\operatorname{aff} \mathcal{C} := \{\theta_1 x_1 + \cdots + \theta_k x_k \mid \forall x_1, \ldots, x_k \in \mathcal{C}, \theta_1 + \cdots + \theta_k = 1\}
$$

- 凸集*(convex set)*

$$
\begin{aligned}
\mathcal{C}\text{为凸集} &\iff \text{过}\mathcal{C}\text{内任意两点的}\textbf{线段}\text{都在}\mathcal{C}\text{内} \\
&\iff \forall x_1, x_2 \in \mathcal{C}, \theta \in \mathbf{[0,1]}, \theta x_1 + (1-\theta)x_2 \in \mathcal{C}
\end{aligned}
$$

- 凸组合

$$
\forall x_1, x_2, \ldots, x_k \in \mathcal{C}, \theta_1, \ldots, \theta_k \in [0,1], \theta_1 + \cdots + \theta_k = 1 : \ \theta_1 x_1 + \cdots + \theta_k x_k \in \mathcal{C}
$$

- 凸包：最小的凸集

$$
\operatorname{conv} \mathcal{C} := \{\theta_1 x_1 + \cdots + \theta_k x_k \mid \forall x_1, \ldots, x_k \in \mathcal{C}, \theta_1, \ldots, \theta_k \in [0,1], \theta_1 + \cdots + \theta_k = 1\}
$$

- 凸锥 *(convex cone)*

$$\mathcal{C}\text{为凸锥} \iff \forall x_1,x_2\in\mathcal{C},\theta_1,\theta_2\geq 0,\theta_1x_1+\theta_2x_2\in\mathcal{C}$$

 除了**空集**的凸锥都得包含**原点**（取$\theta_1=\theta_2=0$）

- 凸锥组合/非负线性组合:

$$\forall x_1,x_2,\ldots,x_k\in\mathcal{C},\theta_1,\ldots,\theta_k\geq 0:\ \theta_1x_1+\cdots+\theta_kx_k\in\mathcal{C}$$

- 凸锥包：类似前面定义

 由上面的定义易知，仿射组合/凸锥组合（强条件）一定是凸组合。

定义 2 (超平面(hyperplane)与半空间(halfspace))**.** 超平面都是比原空间低一维

$$\{\mathbf{x}\mid\mathbf{a}^{\mathrm{T}}\mathbf{x}=b,\mathbf{x},\mathbf{a}\in\mathbb{R}^n,b\in\mathbb{R},\mathbf{a}\neq 0\}$$

超平面将空间划分为两个部分，即半空间

$$\{\mathbf{x}\mid\mathbf{a}^{\mathrm{T}}\mathbf{x}\leq b,\mathbf{a}\neq 0\}$$

若方程特解为$\mathbf{x}_0$，则$\mathbf{a}\perp(\mathbf{x}-\mathbf{x}_0)$

定义 3 (欧式球(Euclidean ball))**.**

$$B(x_c,r)=\{x\mid\|x-x_c\|_2\leq r\}$$

范数*(norm)*球可类似定义

定义 4 (椭球(ellipsoid))**.**

$$\varepsilon(x_c,P)=\{x\mid(x-x_c)^{\mathrm{T}}P^{-1}(x-x_c)\leq 1\},P\succ 0$$

其中$P\succ 0$代表P**对称**且正定*(*$P=P^T$*)*

分析. 定义内积$\langle x^{\mathrm{T}}P^{-1}y\rangle$（需证满足内积条件），进而$P$-范数$\|x\|_P:=\sqrt{x^{\mathrm{T}}Px}$是范数，而椭球不过是$P$-范数意义下的球，由定理得椭球是凸的

定义 5 (多面体(polyhedron))**.**

$$P=\{\mathbf{x}\mid\mathbf{a}_i^{\mathrm{T}}\mathbf{x}\leq b_i,\mathbf{c}_j^{\mathrm{T}}\mathbf{x}=d_j,i=1,\ldots,m,j=1,\ldots,p\}$$

例 3.
- 空集、点、$\mathbb{R}^n$空间均为仿射
- 任意直线为仿射；若过原点则为凸锥

- $\mathbb{R}^n$空间的子空间[3]为仿射和凸锥
- 超平面为仿射
- 半空间、欧式球、椭球、多面体为凸集

定义 6 (仿射函数).

$$f:\mathbb{R}^n \mapsto \mathbb{R}^m \quad f(\mathbf{x}) = A\mathbf{x}+\mathbf{b}, A\in\mathbb{R}^{m\times n}, \mathbf{b}\in\mathbb{R}^m$$

性质如下:

- $S\subset\mathbb{R}^n$为凸 $\implies f(S)=\{f)\mathbf{x}\mid \mathbf{x}\in S\}$为凸
- $C\subset\mathbb{R}^m$为凸 $\implies f^{-1}(C)=\{\mathbf{x}\in\mathbb{R}^n\mid f(\mathbf{x})\in C\}$为凸

例 4. 两个集合的和$S_1+S_2=\{x+y\mid x\in S_1, y\in S_2\}$保凸

分析. 直积$S_1\times S_2=\{(x,y)\mid x\in S_1, y\in S_2\}$显然可以保凸(相当于在两个集合同时画线)
令$A=\begin{bmatrix}I & I\end{bmatrix}, \mathbf{x}=\begin{bmatrix}x & y\end{bmatrix}^{\mathrm{T}}, \mathbf{b}=0$,由仿射函数性质知

定义 7 (透视(perspective)函数[4]). 透视函数$P:\mathbb{R}^{n+1}\mapsto\mathbb{R}^n, \operatorname{dom}P=\mathbb{R}^n\times\mathbb{R}_{++}$定义如下

$$P(z,t)=\frac{z}{t}, z\in\mathbb{R}^n, t\in\mathbb{R}_{++}$$

反透视函数

$$P^{-1}(c):=\{(x,t)\in\mathbb{R}^{n+1}\mid \frac{x}{t}\in c, t>0\}$$

若$c\in\operatorname{dom}P$为凸,则$P(c):=\{P(x), x\in c\}$为凸;反透视函数仍保持c的凸性。

考虑$\mathbb{R}^{n+1}$内的线段,$x=(\widetilde{x}\in\mathbb{R}^n, x_{n+1}\in\mathbb{R}_{++}), y=(\widetilde{y}, y_{n+1})$ 则经过透视函数仍是线段

分析.

$$P(\theta x+(1-\theta)y)=\frac{\theta\widetilde{x}+(1-\theta)\widetilde{y}}{\theta x_{n+1}+(1-\theta y_{n+1})}=\frac{\theta x_{n+1}}{\theta x_{n+1}+(1-\theta)y_{n+1}}\frac{\widetilde{x}}{x_{n+1}}+\frac{(1-\theta)y_{n+1}}{\theta x_{n+1}+(1-\theta)y_{n+1}}\frac{\widetilde{y}}{y_{n+1}}$$

定义 8 (线性分数函数). 仿射函数

$$g(x)=\begin{bmatrix}A\\ C^{\mathrm{T}}\end{bmatrix}x+\begin{bmatrix}b\\ d\end{bmatrix}, A\in\mathbb{R}^{m\times n}, b\in, \mathrm{c}\in\mathbb{R}^{\mathrm{n}}, \mathrm{d}\in\mathbb{R}$$

线性分数函数$f:\mathbb{R}^n\mapsto=\mathrm{p}\circ\mathrm{g}$

$$f(x)=\frac{Ax+b}{c^{\mathrm{T}}x+d}, \operatorname{dom}f=\{x\mid c^{\mathrm{T}}+d>0\}$$

保凸性

- 凸集的交

---

[3]零元、加法封闭、数乘封闭

[4]+代表≥ 0,++代表>0

- 仿射、逆仿射
- 透视函数
- 线性分数函数

3 凸函数

定义 9 (凸函数)**.** *1.* $f:\mathbb{R}^n \to \mathbb{R}$为凸 $\iff$ $\operatorname{dom} f$为凸且$\forall x,y \in \operatorname{dom} f, \theta \in [0,1]$

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

- 严格凸：$\theta \in (0,1)$，不等式不能取等
- 凹函数：若$-f$为凸

2. 高维定义： $f:\mathbb{R}^n \mapsto \mathbb{R}$为凸 $\iff$ $\operatorname{dom} f$为凸

$$\forall x \in \operatorname{dom} f, v \in \mathbb{R}^n : g(t) := f(x+tv)\text{为凸，}\ \operatorname{dom} g = \{t \mid x+tv \in \operatorname{dom} f\}$$

相当于每一个剖面上的低维函数都是凸的

3. 一阶条件*(first-order condition)*[5]

$$f(y) \ge f(x) + \nabla^{\mathrm{T}} f(x)(y-x)$$

例 5. $f(x) = a^{\mathrm{T}}x + b$

分析. 有$\nabla f(x) = a$，进而

$$f(y) = a^{\mathrm{T}}y + b \ge a^{\mathrm{T}}x + b + a^{\mathrm{T}}(y-x) = a^{\mathrm{T}}y + b$$

定义 10 (凸函数的扩展(extended-value))**.** 尽管凸函数的定义域为凸，但往往不好处理，那就将其扩展到全空间。 $x \in \operatorname{dom} f \subset \mathbb{R}^n, \operatorname{dom} \widetilde{f} = \mathbb{R}^n$，会有

$$\widetilde{f}(x) = \begin{cases} f(x) & x \in \operatorname{dom} f \\ +\infty & x \notin \operatorname{dom} f \end{cases}$$

指示/示信(indicator)函数不一定是凸的

$$f(x) = \begin{cases} 0 & x \in C \\ +\infty & x \notin C \end{cases}$$

定理 1. 若f为凸，可微，则$\exists x \in \operatorname{dom} f, \nabla f(x) = 0$

[5] $\nabla^{\mathrm{T}} f(x) = [\nabla f(x)]^{\mathrm{T}}$