凸优化

陆一平 1500010638

2017年2月28日

1 凸集,凸函数,凸优化

1.1 凸集

凸集,仿射集的定义不再叙述,在这里我们先给出几个重要概念的定义与性质:

- **仿射包**: $\operatorname{aff} C = \{\theta_1 x_1 + \cdots + \theta_k x_k | x_1, x_2, \cdots, x_k \in C, \theta_1 + \cdots + \theta_k = 1\}$为包含集合$C$的最小的仿射集合
- **相对内部**:$\operatorname{relint} C = \{x \in C | B(x, r) \cap \operatorname{aff} C \in C, \exists r > 0\}$而相对边界就定义为$\operatorname{cl} C \setminus \operatorname{relint} C$这里$\operatorname{cl} C$表示集合$C$的闭包
- **凸包**:$\operatorname{conv} C = \{\theta_1 x_1 + \cdots + \theta_k x_k | x_1, x_2, \cdots, x_k \in C, \theta_1 + \cdots + \theta_k = 1, \theta_i \geq 0\}$为包含集合$C$的最小的仿射集合

这里容易注意到**交集以及仿射函数都是保凸的运算,**下面我们要展示另外一个不怎么显然的保凸运算:线性分式函数

透视函数

我们定义$P: R^{n+1} \to R^n, P(z, t) = z/t$为透视函数,定义域为$R^n \times R_+$,可以利用小孔成像的方法来解释这个函数也就是通过原点然后投射在了-1这个屏幕上,也是射影平面的一个表示.
通过这个我们可以发现对于凸集C,$P(C)$也是一个凸集,这个结论很直观也可以用数学显示的证明

$$\begin{aligned}
P(\theta x + (1-\theta) y) &= \frac{\theta \tilde{x} + (1-\theta)\tilde{y}}{\theta x_{n+1} + (1-\theta) y_{n+1}} \\
&= \frac{\theta x_{n+1}}{\theta x_{n+1} + (1-\theta) y_{n+1}} P(x) + \frac{(1-\theta) y_{n+1}}{\theta x_{n+1} + (1-\theta) y_{n+1}} P(y)
\end{aligned}$$

类似可以证明一个凸集在透视函数的原像也是凸集
线性分式函数由透视函数以及仿射函数复合而成,对于$f(x) = (Ax+b)/(c^T x + d)$,定义矩阵

$$Q = \begin{bmatrix} A & b \\ c^T & d \end{bmatrix}$$

于是我们有$f(x) = P(QP^{-1}(x))$也是一个保凸函数

1.1.1 广义不等式

锥(conic)定义为一个集合C若$\forall x_1, x_2 \in C, \theta_1, \theta_2 \geq 0, \theta_1 x_1 + \theta_2 x_2 \in C$
一个normal conic定义为一个满足下面条件的锥

- 凸,闭集合
- 有内点
- 是尖的:$x \in K, -x \in K$则有$x = 0$

对于一个normal conic K我们可以定义广义不等式$x \preceq y \Leftrightarrow y - x \in K, x \prec y \Leftrightarrow y - x \in \operatorname{int} K$ 广义不等式对加法保序,具有传递性,对于非负数乘法保序,自反,反对称,对于极限运算保序.在广义不等式下可以定义最小和极小元
对偶锥:对于一个锥K定义对偶锥$K^* = \{y | x^T y \geq 0 \forall x \in K\}$,可以发现对偶锥是一个锥更是一个凸锥. 我们发现对偶锥有如下性质

- K^*是闭凸锥
- 有非空内部的锥的对偶锥是尖的
- 尖的锥的对偶锥有非空内部
- $K^* *$是K的凸包的闭包

于是对于一个广义不等式我们也可以定义他的对偶的不等式,下面我们讨论**最小元的对偶性质**

1.1.2 凸集分离定理

凸集分离定理和支撑超平面定理作为hahn-banach定理的推论这里不再证明与叙述,有兴趣的读者参见张恭庆,林源渠老师的泛函分析讲义(上册)P113.

1.2 凸函数

凸函数的定义不再叙述,在这里给出一条非常有用的凸函数的判定:如果$f : R^n \to R$是凸的,当且仅当函数$g(t) = f(x + tv)$对于每个$v \in R^n$关于t是凸的.

example$f : S^n_{++} \to R, f(X) = logdetX$

Proof.

$$
\begin{aligned}
g(t) &= \log\det(X + tV) = \log\det X + \log\det(I + tX^{-1/2}VX^{1/2}) \\
&= \log\det X + \sum_{i=1}^{n} log(1 + t\lambda_i)
\end{aligned}
$$

where $\lambda_i$are the eigenvalues of $X^{-1/2}VX^{1/2}$
g is concave in t,hence f is concave □

凸函数的epigraph和下水平集都是凸集.同时,凸函数在定义域外可以赋值∞这样可以保持凸性.

凸函数具有一阶与二阶的性质:

- **first order condition:**$f(y) \geq f(x) + f'(x)(y - x)$
- **second order condition:**$\nabla^2 f(x) \succ 0$
- **Jensen's inequality**$f(Ez) \leq Ef(z)$

- **Monotonicity of gradient**$(\nabla f(x)-\nabla f(y))^T(x-y)\geq 0$

  同时凸函数与其epigraph为凸是等价的于是对于关于xw为凸的函数$f(x,y)$则有$g(x)=\sup_y f(x,y)$为凸的

  复合函数的凸性

  对于函数$g:R^n\to R,h:R\to R$,复合函数$f(x)=h(g(x))$在下面任何一种情况下都是凸的

- g凸,h凸,h不减

- g凹,h凸,h不增

这是因为$f''(x)=h''(g(x))g'(x)^2+h'(g(x))g''(x)$
作为推论我们有:$g(x)$凸则$\exp g(x)$凸,$g(x)>0$凸则$1/g(x)$凸

我们不加证明的再给两个凸函数的性质:

- if$f(x,y)$is convex in(x,y) and C is a convex set, then $g(x)=\inf_{y\in C}f(x,y)$is convex

- **the perspective of a function** $f:R^n\to R$ is the function $g:R^n\times R\to R,g(x,t)=tf(x/t)$.If f is convex then g is convex.

有些时候我们还会给凸函数的导数加上李普希兹条件也就是说$||\nabla f(x)-\nabla f(y)||_2\leq L||x-y||_2,\forall x,y$,这时我们有

- $g(x)=(L/2)x^Tx-f(x)$凸

- **Quadratic upper bound:**$f(y)\leq f(x)+\nabla f(x)^T(y-x)+\frac{L}{2}||y-x||_2^2$

如果函数f^*有一个最优解x^*于是

$$\frac{1}{2L}||\nabla f(x)||_2^2\leq f(x)-f(x^*)\leq\frac{L}{2}||x-x^*||^2 \tag{1}$$

右侧是x^*处的二阶上界,而左侧是吧x处的二阶上届给最小化也就是

$$\begin{aligned}f(x^*)&=\min_y(f(x)+\nabla f(x)^T(y-x)+\frac{L}{2}||y-x||_2^2)\\&=f(x)-\frac{1}{2L}||\nabla f(x)||_2^2\end{aligned}$$

强凸函数指的是使得$g(x)=f(x)-\frac{l}{2}x^Tx$为凸函数的$f(x)$,对于强凸函数我们有如下的性质

- **Jensen’s inequality** $f(\theta x+(1-\theta)y)\leq\theta f(x)+(1-\theta)f(y)-\frac{m}{2}\theta(1-\theta)||x-y||_2^2$

- **monotonicity**$(\nabla f(x)-\nabla f(y))^T(x-y)\geq m||x-y||_2^2$

- **first order condition** $f(y)\geq f(x)+f'(x)(y-x)+\frac{m}{2}||y-x||_2^2$

- **second-order condition** $\nabla^2 f(x)\succ mI$

这些性质都是$g(x)$的凸函数的对应性质的等价表示.利用first order condition所给出的二阶的下届我们可以证明

$$\frac{m}{2}\leq f(x)-f(x*)\leq\frac{1}{2m}||\nabla f(x)||_2^2 \tag{2}$$

1.3 凸优化

优化指的是下面的问题

$$\text{minimize} f_0(x) \tag{3}$$
$$\text{subject to} f_i(x) \le 0, i = 1, \cdots, m \tag{4}$$
$$h_i(x) = 0 i = 1, \cdots, p \tag{5}$$
$$x \in C \tag{6}$$

(1)是这个优化问题的目标函数(objective or cost function),(2)是不等式约束(inequality constraint functions),(3)是等式约束,符合约束的x的取值成为可行值(feasible)

最优解(optimal value):$p^* = \inf\{f_0(x)|f_i(x) \le 0, h_i(x) = 0, x \in C\}$

标准的凸优化问题要求目标函数和不等式约束都是凸函数,等式约束应当形如$Ax = b$.也就是说可行集是凸集的凸函数的优化问题.很多时候看起来是非凸规划的问题可以转换为一个凸优化问题进行求解.

example

2 对偶

对于一个凸集我们可以在边界上每个点取支撑超平面,这些支撑超平面可以看作这个凸集合的一种表示,同样我们考虑凸函数,对于每个凸函数,我们用其epigraph的支撑超平面来标出其位置会是怎样的感觉呢?换句话说我们把每个凸函数用一个在其下方的线性函数$< \lambda, x > -\alpha$来逼近这个α要满足$f(x) >< \lambda, x > -\alpha$亦即$\alpha >< \lambda, x > -f(x)$为了取到最佳的$\alpha$我们有

$$\alpha = f^*(\lambda) = \sup_x < \lambda, x > -f(x) \tag{7}$$

这里的函数f^*就称作函数f的对偶函数,我们可以证明f^*一定是一个凸函数

Fenchel’s inequality: $x^T y - f^*(y) \le f(x)$

利用Fenchel's inequality我们可以比较容易的证明$f^{**}(x) \le f(x)$

2.1 次梯度和prox算子

subgradient

如果有

$$f(y) \ge f(x) + g^T(y - x) \forall y$$

g是凸函数f在x处的**次梯度**,于是次梯度给出来一个凸函数的支撑超平面,同时如果一个函数可微,那么梯度也是一个次梯度.次梯度决定了一个函数epigraph的一个支撑超平面.同时也是下水平集的一个支撑超平面.用$\partial f(x)$表示f在x处的次梯度构成的集合.$\partial f(x)$是一个闭的凸集合,对于凸函数$f(x)$在任意一个点的次梯度$\partial f(x)$非空. 对于一个点x,$g \in \partial f(x)$,存在y使得$f(y) \ge f(x) + g^T * (y - x) = f(x) + r||g||_\infty$,在足够小的$r$范围内凸函数是有界$M$的,于是我们有$\sup_{g \in \partial f(x)} ||g||_\infty \le \frac{M - f(x)}{r}$

类似梯度我们可以证明subdifferential of a convex function is a monotone operator,$(u - v)^T(x - y) \ge 0 (\forall u \in \partial f(x), \forall v \in \partial f(y))$

如果$h(x) = f(Ax + b)$则有$\partial h(x) = A^T \partial f(Ax + b)$

Theorem.对于$f(x)=\max\{f_1(x),\cdots,f_n(x)\}$我们定义激活函数$I(x)=\{i|f_i(x)=f(x)\}$，于是有$\partial f(x)=\mathrm{conv}\cup_{i\in I(x)}\partial f_i(x)$

example.计算$f(x)=\lambda_{\max}(A(x))$的一个次梯度,$A(x)=A_0+A_1x_1+\cdots+A_nx_n$

Proof. 找到$\lambda_{\max}(A(x))$的一个特征向量,于是$(y^TA_1y,\cdots,y^TA_ny)\in\partial f(x)$ □

exercise.C是一个凸集,定义$f(x)=\inf_{y\in C}||x-y||_2$的一个次梯度

Theorem.对于闭的凸函数都有

$$y\in\partial f(x)\Leftrightarrow x\in\partial f^*(y)\Leftrightarrow x^Ty=f(x)+f^*(y)$$

Proof. 因为$y\in\partial f(x)$所以有$f^*(y)=\sup_u(y^Tu-f(u))=y^Tx-f(x)$

$$\begin{aligned}f(v)&=\sup_u\{v^Tu-f(u)\}\\&\geq v^Tx-f(x)\\&=x^T(u-y)-f(x)+y^Tx\\&=f^*(y)+x^T(v-y)\end{aligned}$$

□

这个意味着如下一个重要的结论,可以帮助我们通过解决对偶问题来计算原问题

$$\nabla f^*(y)=\arg\min_x(f(x)-x^Ty)\tag{8}$$

proximal mapping

$$\operatorname*{prox}_f(x)=\arg\min_u(f(u)+\frac{1}{2}||u-x||_2^2)\tag{9}$$

同时

$$u=\operatorname*{prox}_f(x)\Leftrightarrow x-u\in\partial f(u)$$

Theorem.$x=\mathrm{prox}_f(x)+\mathrm{prox}_{f^*}x$

Proof.

$$\begin{aligned}u=\operatorname*{prox}_f(x)&\Leftrightarrow x-u\in\partial f(u)\\&\Leftrightarrow u\in\partial f^*(u-x)\\&\Leftrightarrow u-x=\operatorname*{prox}_{f^*}(x)\end{aligned}$$

□

以上的两个结论在后续的应用会是一个重点.我们会非常关心prox_{tf}特别容易计算的f

exercise.计算下列函数的prox_{tf}:二次函数,$f(x)=-\sum_{i=1}^n\log x_i$

exercise.$x=\mathrm{prox}_{\lambda f}(x)+\lambda\,\mathrm{prox}_{\lambda^{-1}f^*}(x/\lambda)$

2.2 凸优化问题的对偶问题

lagrangian dual 对于一个标准形式的凸优化问题

$$\begin{aligned} &\text{minimize} f_0(x) \\ &\text{subject to} f_i(x) \le 0, i=1,2,\cdots,m \\ &\qquad h_i(x)=0, i=1,2,\cdots,p \end{aligned} \qquad x \in \mathcal{D}$$

定义**lagrange函数**$\mathcal{L}(x,\lambda,\mu)=f_0(x)+\sum_{i=1}^m \lambda_i f_i(x)+\sum_{i=1}^p \mu_i h_i(x)$

基于lagrange函数我们可以定义**lagrange 对偶**

$$\begin{aligned} g(\lambda,\mu) &= \inf_{x\in\mathcal{D}}(\mathcal{L}(x,\lambda,\mu)) \\ &= \inf_{x\in\mathcal{D}} f_0(x)+\sum_{i=1}^m \lambda_i f_i(x)+\sum_{i=1}^p \mu_i h_i(x) \end{aligned}$$

对于拉格朗日对偶函数若$\lambda \succeq 0$则有$g(\lambda,\mu) \le p^*$对偶问题就定义为对对偶函数求maximize.而且有弱对偶性质的成立$d^* \le p^*$ 关于拉格朗日对偶函数有多方面的看法,可以参考https://www.zhihu.com/question/48976354中覃含章的答案,或者参考书 [2], [3]. 同时还有讲义https://people.eecs.berkeley.edu/ elghaoui/Teaching/EE227A/lecture7.pd

example.primal problem:minimize $f(x)+g(Ax)$

Proof. 先把原始问题改写

$$\begin{aligned} &\text{minimize } f(x)+g(y) \\ &\text{subject to } y=Ax \end{aligned}$$

Lagrange function:$L(x,y,z)=f(x)+g(y)+z^T(Ax-y)$,
dual function:$\inf_x(f(x)+z^TAx)+\inf_y(g(y)-z^Ty)=-f^*(-Az)-g^*(z)$
dual problem:

$$\text{minimize } f^*(-Az)+g^*(z)$$

□

对偶问题的几何解释

EE 227A: Convex Optimization and Applications	February 14, 2012
Lecture 9: **LP Duality**	
Lecturer: Laurent El Ghaoui	

Reading assignment: Chapter 5 of BV.

9.1 LP Duality

For LPs that are feasible, strong duality always holds.

9.2 Examples

9.2.1 Sum of the largest elements in a vector

For given $x \in \mathbf{R}^n$, and $k \in \{1, \ldots, n\}$ we define

$$s_k(x) := \sum_{i=1}^{k} x_{[i]},$$

where $x_{[i]}$ is the i-th largest element in x.

We can show that

$$s_k(x) = \max_{0 \le u \le \mathbf{1}} \; u^T x \;:\; \sum_{i=1}^{n} u_i = k.$$

Hence the function s_k is convex.

By LP duality, we have

$$s_k(x) = \min_t \; kt + \sum_{i=1}^{n} \max(0, x_i - t).$$

The above representation useful when dealing with a constraint $s_k(x) \le S$ (with S a constant) in an optimization problem, since it allows for a dramatic reduction in the number of constraints needed to represent the constraint.

2.3 优化问题的最优性条件

Complementary slackness 互补松弛条件指的是对于具有强对偶性质的问题,取到最优值的解x^*我们有x^*最小化拉格朗日函数$\mathcal{L}(x,\lambda,\mu)$同时有$\lambda_i f_i(x)=0$,如下的几个条件一起称为最优化问题的**KKT条件**(Karush-Kuhn-Tucker)

- **初始限制条件** $f_i(x)\leq 0, h_i(x)=0$
- **对偶限制** $\lambda \preceq 0$
- **互补松弛**$\lambda_i f_i(x)=0$
- **lagrange函数导数为0**:

$$\nabla f_0(x)+\sum_{i=1}^{m}\lambda_i\nabla f_i(x)+\sum_{i=1}^{p}\mu_i\nabla h_i(x)=0$$

从KKT条件我们知道有$g(\lambda^*,\mu^*)=f_0(x)=L(x^*,\lambda^*,\mu^*)$,可以看到这个条件和拉格朗日乘数法有着异曲同工之妙

3 补充习题1

Homework 3 for "Algorithms for Big-Data Analysis"

Acknowledgement: problems 1, 2, 4 are taken from Prof. Lieven Vandenberghe's course
"EE236C - Optimization Methods for Large-Scale Systems (Spring 2013-14)"

May 5, 2015

1. For each of the following function on $\mathbb{R}^n$, explain how to calculate a subgradient at a given x.
   A reference on subgradients is
   `http://bicmr.pku.edu.cn/~wenzw/bigdata/subgradients.pdf`
   (a) $f(x) = \|Ax - b\|_2 + \|x\|_2$ where $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$.
   (b) $f(x) = \inf_y \|Ay - x\|_\infty$ where $A \in \mathbb{R}^{n\times m}$ and $x \in \mathbb{R}^n$.
2. Give a formula or simple algorithm for evaluating the proximal operator
   $$\mathrm{prox}_f(x) = \arg\min_u \left(f(u) + \frac{1}{2}\|u - x\|_2^2 \right).$$
   (a) $f(x) = \|x\|_1$ with domain $\mathrm{dom}(f) = \{x \mid \|x\|_\infty \le 1\}$
   (b) $f(x) = \max_k x_k$
   (c) $f(x) = \|Ax - b\|_1$ where $AA^\top = D$ and D is a diagonal matrix whose diagonal elements are positive.
3. Given $w \in \mathbb{R}^n$, $\alpha, \sigma > 0$, write down an algorithm for solving the problem
   $$\min_{t,y} \quad \phi(t, y),$$
   where
   $$\phi(t, y) := t + \frac{1}{(1-\alpha)n} \sum_{i=1}^{n} (y_i - t)_+ + \frac{\sigma}{2}\|y - w\|_2^2,$$
   where $x_+ := \max(x, 0)$.
4. Let $S^n = \{X \in \mathbb{R}^{n\times n} \mid X^\top = X\}$ and $S^n_{++} = \{X \in \mathbb{R}^{n\times n} \mid X^\top = X, X \text{ is positive definite }\}$. Find the proximal operator of the function $f(X) = -\log\det X$ where $X \in S^n$ and $\mathrm{dom} f = S^n_{++}$. Here, the proximal operator is defined as
   $$\mathrm{prox}_f(X) = \arg\min_U \left(f(U) + \frac{1}{2}\|U - X\|_F^2 \right),$$
   where $\|\cdot\|_F$ is the Frobenius norm.

1

convex optimization 习题:习题： 4.11, 4.12, 4.25, 4.27, 5.11, 5.17, 5.19

声明:本习题集为文再文老师"大数据中的算法"课程作业

4 基本优化算法

先来列举下优化算法

- gradient method
- subgradient method
- proximal mapping method
- bb method
- cutting-plane methods
- quasi-Newton methods
- conjugate gradient method
- accelerated gradient method
- accelerated proximal mapping method
- ALM method
- Douglas-Rachford splitting method
- ADMM method
- Coordinate Descent Method
- Path-following methods
- Primal-dual interior-point methods

基本的优化算法有线搜索方法与信赖域方法

线搜索方法 理想的线搜索是精确线搜索,但很多时候计算精确线搜索会比较麻烦,我们会有回溯法的搜索方法,回溯的基本框架如下

Algorithm 1 回溯线搜索

Require: r

1: **while** 不满足线搜索条件 **do**

2:     Update $x \leftarrow r * x$

3: **end while**

常用的线搜索准则有:

- **Goldstein准则**

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k g_k^T d_k$$

$$f(x_k + \alpha_k d_k) \geq f(x_k) + (1 - \rho) \alpha_k g_k^T d_k$$

goldstein准则可能把极小点排除在接受区间外侧

- **Wolfe准则**:goldstein准则中第二个条件变为

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k$$

信赖域方法

4.1 梯度下降

梯度下降指的是如下的一个迭代法进行最优化问题的求解

Algorithm 2 梯度下降

---

while not converged **do**

update:

$$x_{k+1} = x_k - t_k \nabla f(x_k) \tag{10}$$

end while

步长t_k的选取颇有技巧有固定步长,bb方法,线搜索方法等

我们先对一个二阶的例子:$f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2)$进行精确的线搜索,我们可以看到下降会走一个崎岖的之字形的路线使得下降速度并没有我们预期的那么快

The performance of an algorithm may depend crucially on how the problem is formulated.Oneimportant issue in problem formulation is scaling. In unconstrained optimization,a problem is said to be poorly scaled if changes to x in a certain direction produce much larger variations in the value of f than do changes to x in another direction.

变量代换解决问题:**diagonal scaling**

Some optimization algorithms, such as steepest descent, are sensitive to poor scaling, while others, such as Newton’s method, are unaffected by it.

Algorithms that are not sensitive to scaling are preferable, because they can handle poor problem formulations in a more robust fashion. In designing complete algorithms, we try to incorporate scale invariance into all aspects of the algorithm, including the line search or trust-region strategies and convergence tests. Generally speaking, it is easier to preserve scale invariance for line search algorithms than for trust-region algorithms.

恒定步长梯度下降的收敛速度

对于可导的凸函数f,我们假设∇f是利普西兹连续的我们考虑如下的迭代方法

$$x_{k+1} = x_k - t_k \nabla f(x_k)$$

我们用$x^+ = x - t\nabla f(x)$利用Quadratic upper bound我们有:

$$\begin{aligned}
f(x^+) &\leq f(x) - \frac{t}{2}||\nabla f(x)||_2^2 (for : t(1 - \frac{Lt}{2}) \geq \frac{t}{2}) \\
&\leq f^* + \nabla f(x)^T (x - x^*) - \frac{t}{2}||\nabla f(x)||_2^2 \\
&= f^* + \frac{1}{2t}(||x - x^*||_2^2 - ||x - x^* - t\nabla f(x)||_2^2) \\
&= f^* + \frac{1}{2t}(||x - x^*||_2^2 - ||x^+ - x^*||_2^2)
\end{aligned}$$

于是我们有:

$$f(x^+) \le f^* + \frac{1}{2t}(||x-x^*||_2^2 - ||x^+ - x^*||_2^2) \tag{11}$$

求和我们有

$$f(x_k) - f(x^*) \le \frac{1}{k}\sum_{i=1}^{k}(f(x_k) - f(x^*)) \le \frac{1}{2kt}||x_0 - x^*||_2^2$$

所以对于步长较小的时候梯度下降是收敛的同时对于误差ϵ我们迭代次数为$O(\frac{1}{\epsilon})$.

下面我们对强凸函数进行分析:

变步长的梯度下降

我们取步长选取为$t_k = \frac{1}{k}$

带有线搜索的梯度下降

我们考虑线搜索准则

$$f(x - t_k\nabla f(x)) \le f(x) - \alpha t_k||\nabla f(x)||_2^2$$

同时更新准则使用$t_k := \beta t_k, 0 < \beta < 1$,为了方便起见我们令$\alpha = \frac{1}{2}$

conclusion: same 1/k bound as with constant step size

Barzilar-Borwein (BB) gradient method

次梯度下降

smoothing

4.2 proximal mapping算法

4.3 Nesterov加速算法

5 Newton法

参考文献

[1] Stephen Boyd Convex Optimization Cambridge University Press

[2] R. TYRRELL ROCKAFELLAR Convex analysis PRINCETON UNIVERSITY PRESS

[3] Dimitri P.Berteskas Convex Optimization Theory

[4] Dimitri P.Berteskas Convex Optimization Algorithms

[5] A Beck，M Teboulle A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems Siam Journal on Imaging Sciences

[6] Berkeley EE227A

[7] Dimitri Bertsekas. 6.253 Convex Analysis and Optimization. Spring 2012. Massachusetts Institute of Technology: MIT OpenCourseWare
Stephen Boyd, and Pablo Parrilo. 6.079 Introduction to Convex Optimization. Fall 2009. Massachusetts Institute of Technology: MIT OpenCourseWare

[8] “Numerical Optimization”, Jorge Nocedal and Stephen Wright, Springer