Convex Optimization Theory and Applications

Topic 21 - Others

Li Li

Department of Automation
Tsinghua University

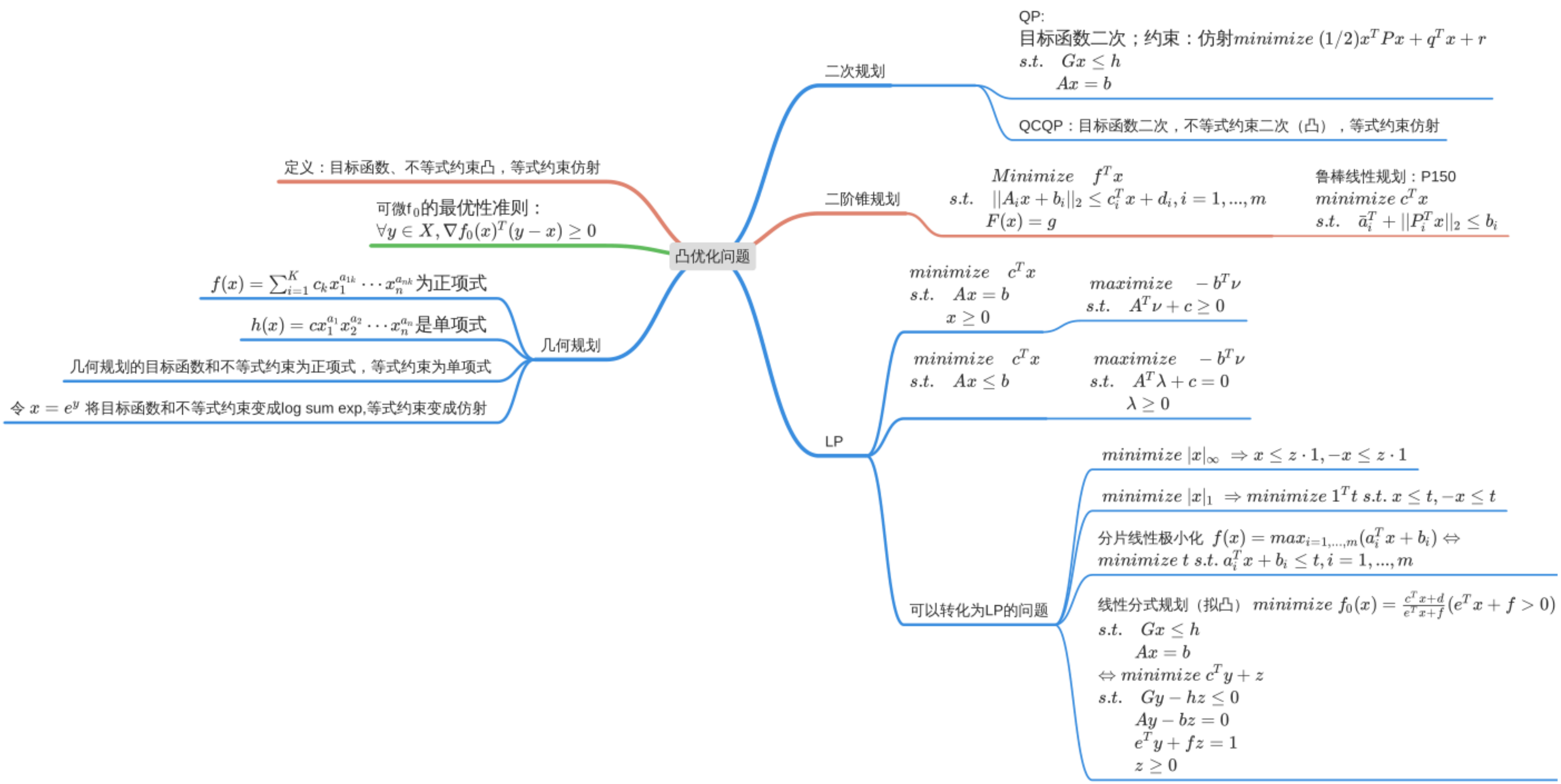
Fall, 2009-2021.

21.0. Outline

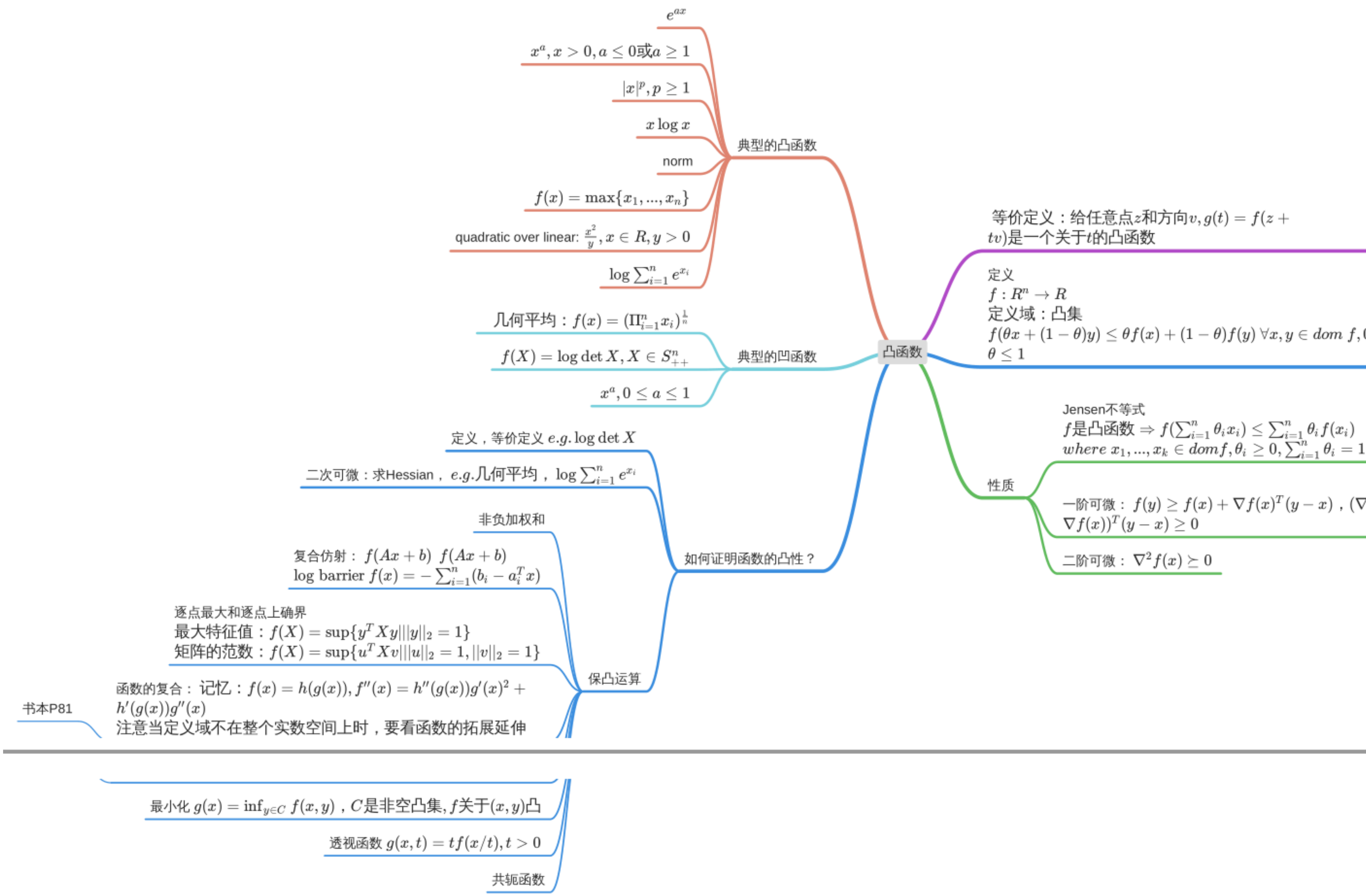
21.1. Mind Map

欢迎大家使用思维导图来帮助记忆和整理

2020 年顾欣然同学的思维导图



21.1. Mind Map



凸函数
典型的凸函数
e^{ax}
$x^a, x>0, a\le 0$或$a\ge 1$
$|x|^p, p\ge 1$
$x\log x$
norm
$f(x)=\max\{x_1,...,x_n\}$
quadratic over linear: $\frac{x^2}{y}, x\in R, y>0$
$\log\sum_{i=1}^n e^{x_i}$
典型的凹函数
几何平均：$f(x)=(\Pi_{i=1}^n x_i)^{\frac{1}{n}}$
$f(X)=\log\det X, X\in S_{++}^n$
$x^a, 0\le a\le 1$
如何证明函数的凸性？
定义，等价定义 $e.g. \log\det X$
二次可微：求Hessian， $e.g.$几何平均， $\log\sum_{i=1}^n e^{x_i}$
保凸运算
非负加权和
复合仿射： $f(Ax+b)$ $f(Ax+b)$
log barrier $f(x)=-\sum_{i=1}^n(b_i-a_i^Tx)$
逐点最大和逐点上确界
最大特征值：$f(X)=\sup\{y^TXy\,|\,\|y\|_2=1\}$
矩阵的范数：$f(X)=\sup\{u^TXv\,|\,\|u\|_2=1,\|v\|_2=1\}$
函数的复合： 记忆：$f(x)=h(g(x)), f''(x)=h''(g(x))g'(x)^2+h'(g(x))g''(x)$
注意当定义域不在整个实数空间上时，要看函数的拓展延伸
书本P81
最小化 $g(x)=\inf_{y\in C} f(x,y)$，$C$是非空凸集, f关于(x,y)凸
透视函数 $g(x,t)=tf(x/t), t>0$
共轭函数
等价定义：给任意点z和方向v, $g(t)=f(z+tv)$是一个关于t的凸函数
定义
$f: R^n\to R$
定义域：凸集
$f(\theta x+(1-\theta)y)\le\theta f(x)+(1-\theta)f(y)\ \forall x,y\in dom\ f, 0$
$\theta\le 1$
性质
Jensen不等式
f是凸函数 $\Rightarrow f(\sum_{i=1}^n\theta_i x_i)\le\sum_{i=1}^n\theta_i f(x_i)$
$where\ x_1,...,x_k\in dom f, \theta_i\ge 0, \sum_{i=1}^n\theta_i=1$
一阶可微： $f(y)\ge f(x)+\nabla f(x)^T(y-x)$，$(\nabla$
$\nabla f(x))^T(y-x)\ge 0$
二阶可微： $\nabla^2 f(x)\succeq 0$

21.1. Mind Map

"A Mind Map is a diagram for representing tasks, words, concepts, or items linked to and arranged around a central concept or subject using a non-linear graphical layout that allows the user to build an intuitive framework around a central concept.

In California, Dr. Roger Sperry, who won a Nobel Prize for his research, confirmed that the evolutionarily latest part of the brain, the 'thinking cap' of the Cerebral Cortex, was divided into two major hemispheres, and those hemispheres performed a comprehensive range of intellectual tasks, called cortical skills. The tasks included: Logic, Rhythm, Lines, Color, Lists, Daydreaming, Numbers, Imagination, Word, Gestalt (seeing the whole picture)."

https://www.mindmapping.com/mind-map

21.2. First

2.12 Which of the following sets are convex?

(a) A *slab*, *i.e.*, a set of the form $\{x \in \mathbf{R}^n \mid \alpha \leq a^T x \leq \beta\}$.

(b) A *rectangle*, *i.e.*, a set of the form $\{x \in \mathbf{R}^n \mid \alpha_i \leq x_i \leq \beta_i,\ i = 1, \ldots, n\}$. A rectangle is sometimes called a *hyperrectangle* when $n > 2$.

(c) A *wedge*, *i.e.*, $\{x \in \mathbf{R}^n \mid a_1^T x \leq b_1,\ a_2^T x \leq b_2\}$.

(d) The set of points closer to a given point than a given set, *i.e.*,

$$\{x \mid \|x - x_0\|_2 \leq \|x - y\|_2 \text{ for all } y \in S\}$$

where $S \subseteq \mathbf{R}^n$.

(e) The set of points closer to one set than another, *i.e.*,

$$\{x \mid \mathbf{dist}(x, S) \leq \mathbf{dist}(x, T)\},$$

where $S, T \subseteq \mathbf{R}^n$, and

$$\mathbf{dist}(x, S) = \inf\{\|x - z\|_2 \mid z \in S\}.$$

(f) [HUL93, volume 1, page 93] The set $\{x \mid x + S_2 \subseteq S_1\}$, where $S_1, S_2 \subseteq \mathbf{R}^n$ with S_1 convex.

(g) The set of points whose distance to a does not exceed a fixed fraction θ of the distance to b, *i.e.*, the set $\{x \mid \|x - a\|_2 \leq \theta \|x - b\|_2\}$. You can assume $a \neq b$ and $0 \leq \theta \leq 1$.

21.2. First

(a) A slab is an intersection of two halfspaces, hence it is a convex set and a polyhedron.

(b) As in part (a), a rectangle is a convex set and a polyhedron because it is a finite intersection of halfspaces.

(c) A wedge is an intersection of two halfspaces, so it is convex and a polyhedron. It is a cone if $b_1 = 0$ and $b_2 = 0$.

(d) This set is convex because it can be expressed as

$$\bigcap_{y \in S} \{x \mid \|x - x_0\|_2 \leq \|x - y\|_2\},$$

i.e., an intersection of halfspaces. (Recall from exercise 2.9 that, for fixed y, the set

$$\{x \mid \|x - x_0\|_2 \leq \|x - y\|_2\}$$

is a halfspace.)

(e) In general this set is not convex, as the following example in $\mathbf{R}$ shows. With $S = \{-1, 1\}$ and $T = \{0\}$, we have

$$\{x \mid \mathbf{dist}(x, S) \leq \mathbf{dist}(x, T)\} = \{x \in \mathbf{R} \mid x \leq -1/2 \text{ or } x \geq 1/2\}$$

which clearly is not convex.

21.2. First

(f) This set is convex. $x + S_2 \subseteq S_1$ if $x + y \in S_1$ for all $y \in S_2$. Therefore

$$\{x \mid x + S_2 \subseteq S_1\} = \bigcap_{y \in S_2} \{x \mid x + y \in S_1\} = \bigcap_{y \in S_2} (S_1 - y),$$

the intersection of convex sets $S_1 - y$.

(g) The set is convex, in fact a ball.

$$\begin{aligned} &\{x \mid \|x - a\|_2 \leq \theta \|x - b\|_2\} \\ &= \quad \{x \mid \|x - a\|_2^2 \leq \theta^2 \|x - b\|_2^2\} \\ &= \quad \{x \mid (1 - \theta^2) x^T x - 2(a - \theta^2 b)^T x + (a^T a - \theta^2 b^T b) \leq 0\} \end{aligned}$$

If $\theta = 1$, this is a halfspace. If $\theta < 1$, it is a ball

$$\{x \mid (x - x_0)^T (x - x_0) \leq R^2\},$$

with center x_0 and radius R given by

$$x_0 = \frac{a - \theta^2 b}{1 - \theta^2}, \qquad R = \left(\frac{\theta^2 \|b\|_2^2 - \|a\|_2^2}{1 - \theta^2} - \|x_0\|_2^2 \right)^{1/2}.$$

21.2. First

2.9 *Voronoi sets and polyhedral decomposition.* Let $x_0, \ldots, x_K \in \mathbf{R}^n$. Consider the set of points that are closer (in Euclidean norm) to x_0 than the other x_i, *i.e.*,

$$V = \{x \in \mathbf{R}^n \mid \|x - x_0\|_2 \leq \|x - x_i\|_2, \ i = 1, \ldots, K\}.$$

V is called the *Voronoi region* around x_0 with respect to $x_1, \ldots, x_K$.

(a) Show that V is a polyhedron. Express V in the form $V = \{x \mid Ax \preceq b\}$.

(b) Conversely, given a polyhedron P with nonempty interior, show how to find $x_0, \ldots, x_K$ so that the polyhedron is the Voronoi region of x_0 with respect to $x_1, \ldots, x_K$.

(c) We can also consider the sets

$$V_k = \{x \in \mathbf{R}^n \mid \|x - x_k\|_2 \leq \|x - x_i\|_2, \ i \neq k\}.$$

The set V_k consists of points in $\mathbf{R}^n$ for which the closest point in the set $\{x_0, \ldots, x_K\}$ is x_k.

The sets $V_0, \ldots, V_K$ give a polyhedral decomposition of $\mathbf{R}^n$. More precisely, the sets V_k are polyhedra, $\bigcup_{k=0}^{K} V_k = \mathbf{R}^n$, and $\mathbf{int}\, V_i \cap \mathbf{int}\, V_j = \emptyset$ for $i \neq j$, *i.e.*, V_i and V_j intersect at most along a boundary.

Suppose that $P_1, \ldots, P_m$ are polyhedra such that $\bigcup_{i=1}^{m} P_i = \mathbf{R}^n$, and $\mathbf{int}\, P_i \cap \mathbf{int}\, P_j = \emptyset$ for $i \neq j$. Can this polyhedral decomposition of $\mathbf{R}^n$ be described as the Voronoi regions generated by an appropriate set of points?

21.2. First

2.9 *Voronoi sets and polyhedral decomposition.* Let $x_0, \ldots, x_K \in \mathbf{R}^n$. Consider the set of points that are closer (in Euclidean norm) to x_0 than the other x_i, *i.e.*,

$$V = \{x \in \mathbf{R}^n \mid \|x - x_0\|_2 \leq \|x - x_i\|_2, \; i = 1, \ldots, K\}.$$

V is called the *Voronoi region* around x_0 with respect to $x_1, \ldots, x_K$.

(a) Show that V is a polyhedron. Express V in the form $V = \{x \mid Ax \preceq b\}$.

(b) Conversely, given a polyhedron P with nonempty interior, show how to find $x_0, \ldots, x_K$ so that the polyhedron is the Voronoi region of x_0 with respect to $x_1, \ldots, x_K$.

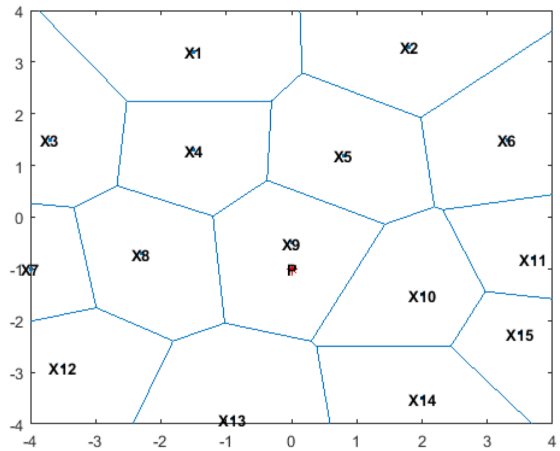
(c) We can also consider the sets

$$V_k = \{x \in \mathbf{R}^n \mid \|x - x_k\|_2 \leq \|x - x_i\|_2, \; i \neq k\}.$$

The set V_k consists of points in $\mathbf{R}^n$ for which the closest point in the set $\{x_0, \ldots, x_K\}$ is x_k.

The sets $V_0, \ldots, V_K$ give a polyhedral decomposition of $\mathbf{R}^n$. More precisely, the sets V_k are polyhedra, $\bigcup_{k=0}^{K} V_k = \mathbf{R}^n$, and $\textbf{int}\, V_i \cap \textbf{int}\, V_j = \emptyset$ for $i \neq j$, *i.e.*, V_i and V_j intersect at most along a boundary.
Suppose that $P_1, \ldots, P_m$ are polyhedra such that $\bigcup_{i=1}^{m} P_i = \mathbf{R}^n$, and $\textbf{int}\, P_i \cap \textbf{int}\, P_j = \emptyset$ for $i \neq j$. Can this polyhedral decomposition of $\mathbf{R}^n$ be described as the Voronoi regions generated by an appropriate set of points?

21.2. First




```
figure()
X = [-1.5 3.2; 1.8 3.3; -3.7 1.5; -1.5 1.3; 0.8 1.2; 3.3 1.5; -4.0 -1.0;-2.3 -0.7; 0 -0.5; 2.0 -1.5;
3.7 -0.8; -3.5 -2.9; -0.9 -3.9; 2.0 -3.5; 3.5 -2.25];

voronoi(X(:,1),X(:,2))

% Assign labels to the points.
nump = size(X,1);
plabels = arrayfun(@(n) {sprintf('X%d', n)}, (1:nump)');
hold on
Hpl = text(X(:,1), X(:,2), plabels, 'FontWeight', 'bold', 'HorizontalAlignment','center',
'BackgroundColor', 'none');
```

21.3. Second

Please use the defition of convex function to prove the convexity of conjugate function

21.3. Second

Please use the defition of convex function to prove the convexity of conjugate function

> $\forall x \in \mathrm{dom}(f)$, is $h(y,\, x) := y^T x - f(x)$ convex in y?

Let us try to prove this using Jensen's inequality.

For some $x,\, a,\, b \in \mathrm{dom}(f),\, \theta \in [0,\, 1]$ consider:

$$\begin{aligned} & h(\theta a + (1-\theta)b,\, x) \\ & = (\theta a + (1-\theta)b)^T x - f(x) \\ & = \theta a^T x + (1-\theta)b^T x - f(x) \\ & = \theta a^T x - \theta f(x) + (1-\theta)b^T x - (1-\theta)f(x) \\ & = \theta(a^T x - f(x)) + (1-\theta)(b^T x - f(x)) \\ & = \theta h(a,\, x) + (1-\theta)h(b,\, x) \end{aligned}$$

Since equality always holds, $h(y,\, x)$ is convex in y.

Now that I have done this, I realize the following: $h(y,\, x)$ was just an affine function in y, meaning that the pointwise supremum, *i.e.* the conjugate of f, will also be convex.

21.3. Second

the **conjugate** of a function f is

$$f^*(y) = \sup_{x \in \mathrm{dom}\, f} (y^T x - f(x))$$

Fenchel's inequality: the definition implies that

$$f(x) + f^*(y) \geq x^T y \quad \text{for all } x, y$$

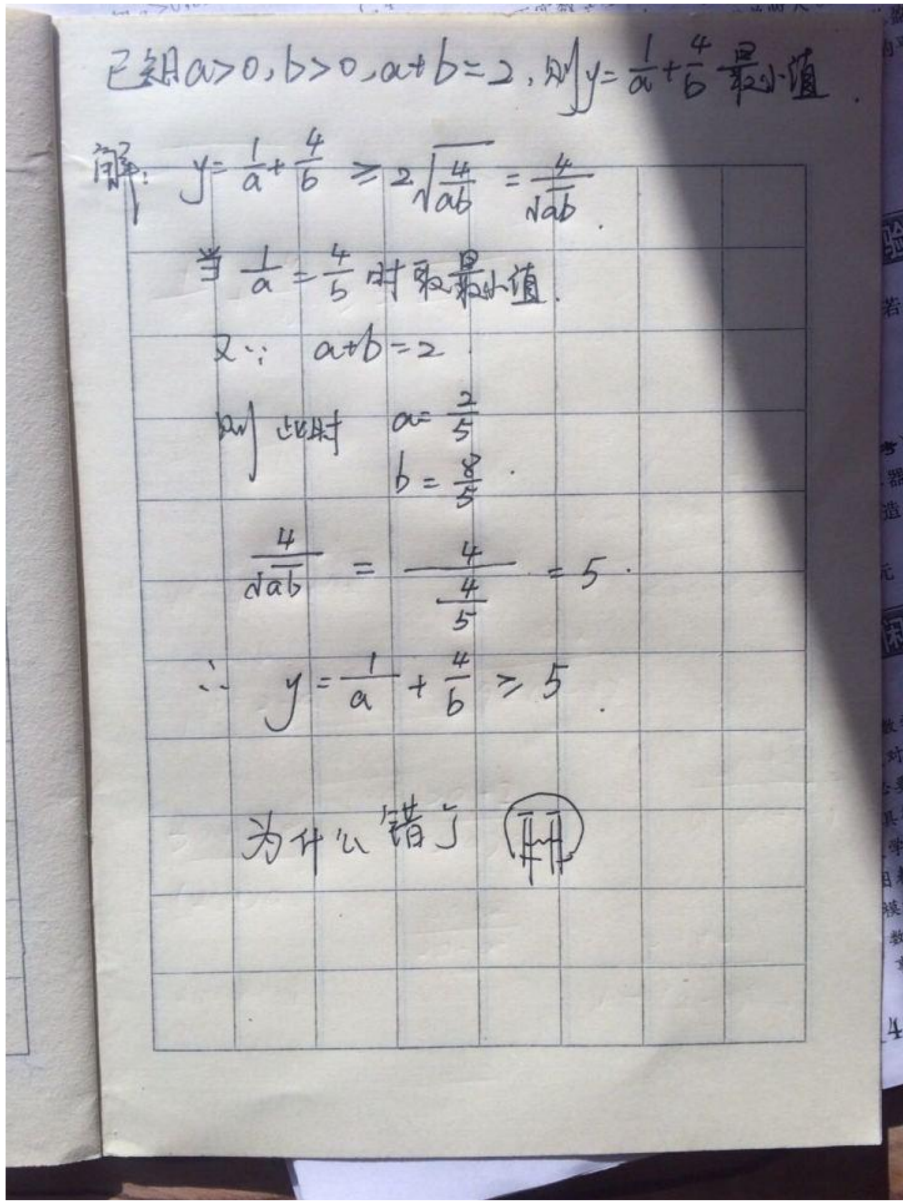
this is an extension to non-quadratic convex f of the inequality

$$\frac{1}{2} x^T x + \frac{1}{2} y^T y \geq x^T y$$

21.3. Second

请证明 l_1，l_∞ 范数小于定值的集合是多面体的

21.3. Second



已知$a>0, b>0, a+b=2$，则$y=\frac{1}{a}+\frac{4}{b}$最小值.

解：$y=\frac{1}{a}+\frac{4}{b}\geq 2\sqrt{\frac{4}{ab}}=\frac{4}{\sqrt{ab}}$.

当$\frac{1}{a}=\frac{4}{b}$时取最小值.

又∵ $a+b=2$

则此时 $a=\frac{2}{5}$

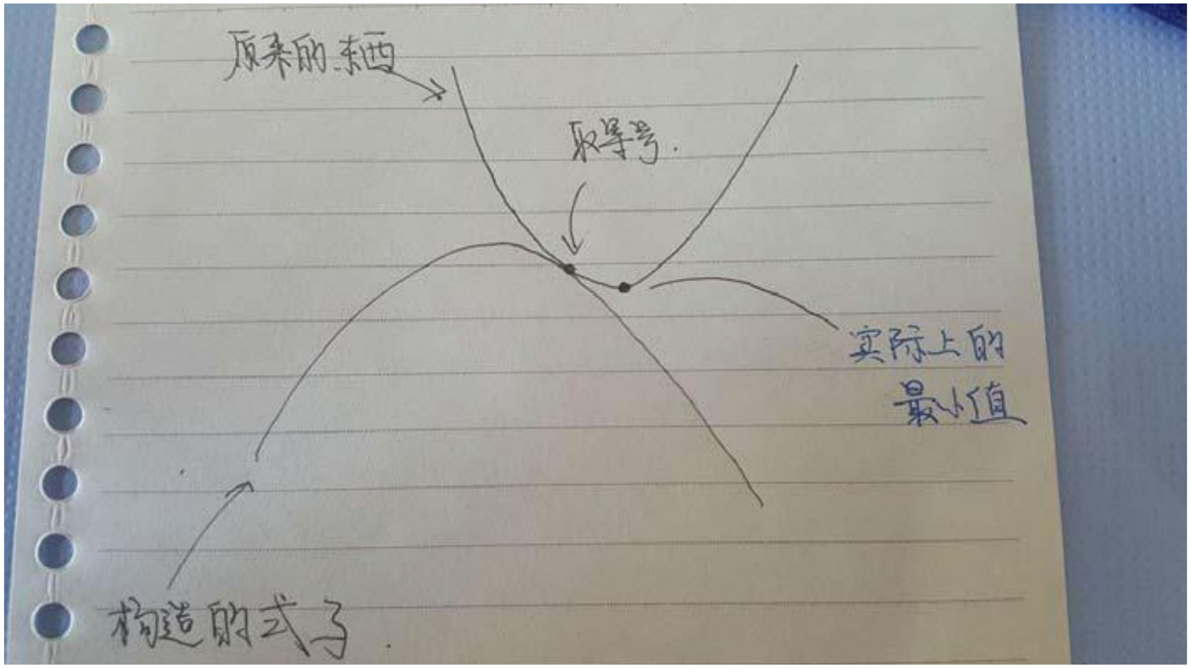
$b=\frac{8}{5}$.

$$\frac{4}{\sqrt{ab}}=\frac{4}{\frac{4}{5}}=5.$$

∴ $y=\frac{1}{a}+\frac{4}{b}\geq 5$.

为什么错了

21.3. Second



$y=\frac{1}{a}+\frac{4}{b} \geq \frac{4}{\sqrt{ab}}$

只能推出 $\frac{1}{a}=\frac{4}{b}$ 时 $y=\frac{4}{\sqrt{ab}}$

而 $\frac{4}{\sqrt{ab}}$ 是一个变化的值

∴会有一种情况 $\frac{1}{a}\neq\frac{4}{b}$ $y>\frac{4}{\sqrt{ab}}$ 但此时 $\sqrt{ab}$ 比 $\frac{1}{a}=\frac{4}{b}$ 时的 $\sqrt{ab}$ 更大 所以 y 可能更小

21.3. Second

还是Jensen's Inequality

$$(a+b)\left(\frac{1}{a}+\frac{4}{b}\right) \geq \left(\sqrt{a\cdot\frac{1}{a}}+\sqrt{b\cdot\frac{4}{b}}\right) = 9$$

21.3. Second

东京大学1999年高考第六题(理科)

Please prove $\int_0^{\pi} e^x \sin^2 x dx > 8$

21.3. Second

$$\int e^x sin^2 x dx = \int (e^x)' sin^2 x dx = e^x sin^2 x dx - \int (e^x)' cos x \times 2 sin x dx$$

$$= e^x sin^2 x dx - \int e^x sin 2x dx$$

下面我们来求 $\int e^x sin2xdx$ 是多少

$$(e^x cos2x)' = -2e^x sin2x + e^x cos2x......(1)$$

$$(e^x sin2x)' = 2e^x cos2x + e^x sin2x......(2)$$

$$(2) - 2 \times (1):$$

$$(-2e^x cos2x + e^x sin2x)' = 5e^x sin2x$$

21.3. Second

因此，解得

$$\int e^x sin2xdx = \frac{e^x sin2x - 2e^x cos2x}{5}$$

所以

$$\int e^x sin^2 xdx = e^x sin^2 xdx - \frac{e^x sin2x - 2e^x cos2x}{5}$$

带入积分上下限

$$\int_0^\pi e^x sin^2 xdx = (0 - \frac{0 - 2e^\pi}{5}) - (0 - \frac{0-2}{5}) = \frac{2e^\pi - 2}{5}$$

一通化简之后，Please prove $e^\pi > 21$

21.3. Second

考虑到e^x是凸函数，做$x=3$这个点处e^x的切线分析

在本题中，为了方便计算，我们设 $t=3$ ，那么我们想要求的
$x=\pi$ 处的值有如下关系

$$e^{\pi}-g(\pi)>0$$

函数 $g(x)$ 是 e^x 在$(3,e^3)$ 处引出的切线，斜率C为 e^3

因此 $g(x)$ 的表达式为 $g(x)=e^3(x-3)+e^3$

$x=\pi$ 时，可得

$$e^{\pi}-e^3\times(\pi-3)-e^3>0$$

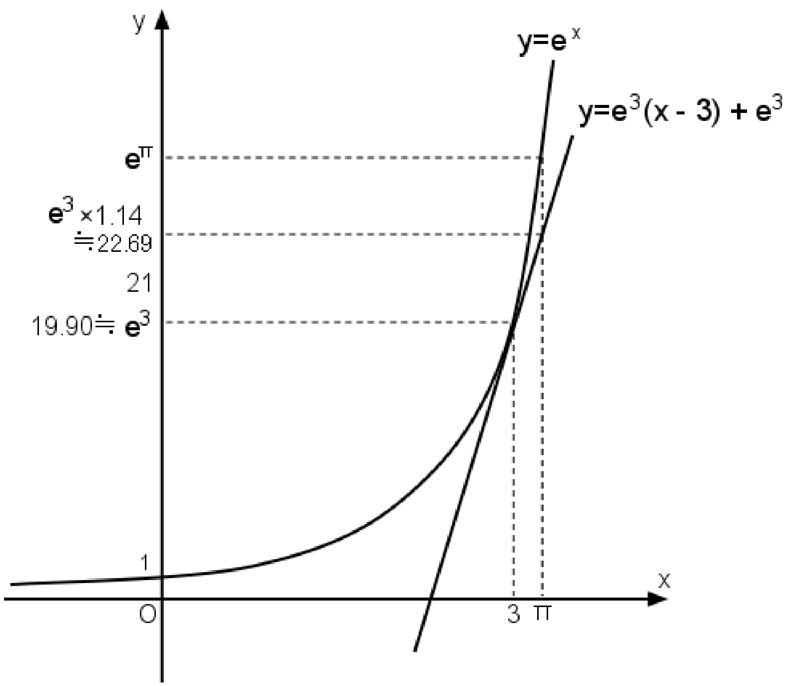
$$e^{\pi}>\pi e^3-2e^3$$

在这里就可以放缩了， $\pi=3.1,e=2.7$

21.3. Second

$$e^{\pi} > (3.1-2)\times 2.7^{3} = 1.1\times 2.7^{3} = 21.6513$$

因此 $e^{\pi} > 21$ 成立



21.4. Third

Gradient descent on a convex function with Lipschitz gradient converges (with suitable step sizes) at the rate:
☐ a. $O(1/\epsilon^2)$;
☐ b. $O(1/\epsilon)$;
☐ c. $O(\log(1/\epsilon))$;
☐ d. $O(1/\sqrt{\epsilon})$.

Consider the backtracking loop:

$$\text{repeat} \quad t = \beta t$$
$$\text{until} \quad f(x + tv) \leq f(x) + \alpha t \nabla f(x)^T v,$$

where v is the descent direction. A larger value of α leads to fewer backtracking iterations until the backtracking exit condition is satisfied.
☐ True
☐ False

Applying Nesterov's acceleration to gradient descent still results in a descent method.
☐ True
☐ False

21.4. Third

Consider Newton's method on a convex function, under lower and upper bounds on the eigenvalues of the Hessian, and a Lipschitz Hessian. In its region of quadratic convergence, backtracking will always select pure step sizes (equal to 1).
☐ True
☐ False

Which of the following statements comparing Newton's method and gradient descent is not accurate (for convex problems)?
☐ a. One Newton iteration is typically more computationally costly than one gradient descent iteration.
☐ b. Newton's method has faster local rate of convergence under suitable assumptions.
☐ c. Each Newton step can be viewed as an exact minimization of a suitable quadratic approximation, whereas that is not the case for gradient descent.
☐ d. In both Newton's method and gradient descent, we can use backtracking to ensure global convergence.

Generally speaking, Newton's method can be used for: a. finding the roots of a nonlinear equation, b. minimizing a function. These ideas are completely separate (they don't have any relationship).
☐ True
☐ False

Compared to gradient descent, Newton's method, roughly speaking:
☐ a. uses more accurate quadratic approximations, admits more expensive iterations, but requires fewer iterations to converge to high accuracy;
☐ b. uses less accurate quadratic approximations, and cheaper iterations, so it requires more iterations to converge to high accuracy;
☐ c. uses cubic approximations, and its iterations and convergence are not really comparable to gradient descent;
☐ d. only approximates the smooth part of the criterion by a quadratic, and thus applies to a broader class of nonsmooth optimization problems.

21.4. Third

If we run Newton's method on $f(x)$ starting at x_0 for k iterations, and for the same sequence of step sizes, run Newton's method on $g(y) = f(Ay)$ starting at $y_0 = A^{-1}x_0$ for k iterations, then the achieved sequence of criterion values is the same in both cases.
☐ True
☐ False

Pure Newton's method (with step sizes equal to 1) will always converge on a convex function.
☐ True
☐ False

Newton's method on a convex function, under lower and upper bounds on the eigenvalues of the Hessian, and a Lipschitz Hessian, converges (with suitable step sizes) at the rate:
☐ a. $O(\log\log(1/\epsilon))$;
☐ b. $O(\log\log(1/\epsilon))$, locally;
☐ c. it depends on whether the function is self-concordant;
☐ d. none of the above.

Let $f^\star$ denote the optimal criterion value of the convex problem

$$\min_x \ f(x) \ \text{ subject to } \ h_j(x) \le 0, \ j = 1, \ldots, m,$$

and $x^\star(t)$ denote the solution in the barrier problem

$$\min_x \ tf(x) + \phi(x).$$

Then $f(x^\star(t)) - f^\star \le m/t$.
☐ True
☐ False

21.4. Third

Each main iteration of the barrier method performs just one one Newton update.
☐ True
☐ False

The main idea behind the barrier method is add terms to the criterion that:
☐ a. smoothly approximate indicator functions of the constraints;
☐ b. make the new criterion strongly convex;
☐ c. make the new criterion smooth;
☐ d. get rid of equality constraints.

The barrier method solves the problem:

$$\min_x \; f(x) \;\; \text{subject to} \;\; h_j(x) \leq 0, \; j = 1, \ldots, m,$$

by solving a:
☐ a. single problem of the form $\min_x (tf(x) + \phi(x))$, where $\phi(x) = -\sum_{j=1}^m \log(-h_j(x))$ and $t > 0$;
☐ b. sequence of problems of the form $\min_x (t_k f(x) + \phi(x))$, where $\phi(x) = -\sum_{j=1}^m \log(-h_j(x))$ and $t_k \to \infty$;
☐ c. sequence of problems of the form $\min_x (t_k f(x) + \phi(x))$, where $\phi(x) = -\sum_{j=1}^m \log(-h_j(x))$ and $t_k \to 0$;
☐ d. sequence of problems of the form $\min_x (t_k f(x) + \phi(x))$, where $\phi(x) = \sum_{j=1}^m \log(-h_j(x))$ and $t_k \to \infty$;
☐ e. sequence of problems of the form $\min_x (t_k f(x) + \phi(x))$, where $\phi(x) = \sum_{j=1}^m \log(-h_j(x))$ and $t_k \to 0$.

For constrained convex minimization, barrier methods approach the solution from the outside of the constraint set.

☐ True

☐ False

Which of the following statements about the barrier method and the primal-dual interior-point method is not true (for convex problems)?

☐ a. Both barrier method and primal-dual interior-point method can be interpreted as solving a perturbed version of the KKT conditions.

☐ b. Both methods have local $O(\log(1/\epsilon))$ rate of convergence.

☐ c. Primal-dual interior-point method is more commonly used in practice because it tends to be more efficient.

☐ d. Both methods perform just one Newton update before taking a step along the central path (adjusting the barrier parameter t).

The iterates of the primal-dual interior-point method are always primal and dual feasible.

☐ True

☐ False

Each main iteration of a primal-dual interior-point method performs just one one Newton update.

☐ True

☐ False

Which of the following statements about the barrier method and the primal-dual interior-point method is not true (for convex problems)?

☐ a. Both barrier method and primal-dual interior-point method can be interpreted as solving a perturbed version of the KKT conditions.

☐ b. Both require solving a linear system at the lowest level of iteration.

☐ c. Both methods have local $O(\log(1/\epsilon))$ rate of convergence.

☐ d. Both yield feasible primal and dual iterates at every step.

21.4. Third

Consider the convex problem:

$$\min_x f(x) \text{ subject to } Ax = b,\ h(x) \leq 0.$$

Which one of the following is not a consideration, when choosing the step size in each main iteration of a primal-dual interior-point algorithm applied to the this problem?

☐ a. Take a full Newton step (step size equal to 1) if possible.
☐ b. Take a step size that ensures $h(x) < 0$.
☐ c. Take a step size that ensures $u > 0$.
☐ d. Take a step size that ensures $Ax = b$.

21.5. Fourth

1. （10分）设K为正常锥，$A\in R^{m\times n}$，$c\in R^n$。记K的对偶锥为K^*，假设A^TK^*为闭集。请证明：$Ax\succeq_K 0,\ c^Tx<0$无解等价于$c=A^Ty,\ y\succeq_{K^*}0$有解。其中$\succeq_K$和$\succeq_{K^*}$是广义不等式。

2. （10分）判断以下函数$f(\cdot)$在相应$\mathbf{dom}\ f$上的凹凸性，并给出理由。

   1） $f(X)=\log\det X$，$\mathbf{dom}\,f=S_{++}^n$

   2） $f(x)=\sum_{i=1}^{r}|x|_{[i]}$，$\mathbf{dom}\,f=R^n$，其中$1\le r\le n$，$|x|_{[i]}$表示$x$的分量的绝对值中第$i$大数，即，对任意$x\in R^n$成立$|x|_{[1]}\ge|x|_{[2]}\ge\cdots\ge|x|_{[n]}$

21.5. Fourth

3. （20分）对以下两个优化问题，请说明是否可以将其转化为凸优化问题，并给出理由。

1) $\min\left\{\dfrac{\|Ax-b\|_1}{c^Tx+d}\ \middle|\ \text{s.t. } \|x\|_\infty\le 1\right\}$，其中$c\in R^n$，$d\in R$，$d>\|c\|_1$；

2)
$$\begin{aligned}
\min\ & \frac{f_1(x)}{f_2(x)-f_3(x)}\\
\text{s.t. } & f_1(x)=\sum_{j=1}^{4} j\, x_1^{\alpha_{j1}}x_2^{\alpha_{j2}}x_3^{\alpha_{j3}}\\
& f_2(x)=\sum_{j=1}^{4} j^2\, x_1^{\beta_{j1}}x_2^{\beta_{j2}}x_3^{\beta_{j3}}\\
& f_3(x)=x_1x_2x_3\\
& f_2(x)>f_3(x)\\
& x_i\ge\varepsilon,\ i=1,2,3
\end{aligned}$$

其中α_{ji}，β_{ji}，ε是给定实数，$\varepsilon>0$。

21.5. Fourth

4. （20分）求解以下凸优化问题的最优解。（提示：利用KKT条件，对于互补松弛条件进行讨论）

$$\min \left\{-\log(x+1)-\log(y+2) \,\middle|\, \text{s.t. } x \geq 0, y \geq 0, x+y=3\right\}$$

6. 对于如下优化问题：

$$\begin{aligned} \min_x \quad & x^T P x \\ s.t. \quad & Ax = b \end{aligned}$$

其中 $x \in R^n$, $P \in S_{++}^n$, $A \in R^{m \times n}$ 。

请解释它的几何意义，并推导其对偶问题。

21.5. Fourth

5. （10分）考虑线性规划问题$\min\{\ 5x_1+11x_2 \mid \text{s.t. } x\in\Omega\}$，其中

$$\Omega=\{\ x\in R^2 \mid 6x_1+x_2\ge 6,\ 5x_1+2x_2\ge 10,\ 4x_1+3x_2\ge 12$$
$$3x_1+4x_2\ge 12,\ 2x_1+5x_2\ge 10,\ x_1+6x_2\ge 6\ \}$$

1) 用对数障碍函数处理不等式约束，设定障碍函数的惩罚参数$t=1$，将问题转化为无约束问题。
2) 用梯度下降法求解相应的无约束问题，用回溯直线搜索方法进行一维搜索，回溯直线搜索参数为$\alpha=0.2$，$\beta=0.6$。请写出以$x_1=x_2=2$为初值的一步迭代公式，包括直线搜索方向、直线搜索的目标函数以及直线搜索的停止准则。要求所写出的公式均是已知数据的表达式，但不需要计算具体数值完成以下工作。

21.5. Fourth

要求：请写出详细的推导过程或证明过程。

1. （20分）证明集合$S_+^n = \{X \in \mathbb{R}^{n\times n} \mid X^\top = X, \quad X \succeq 0\}$为凸锥(convex cone)，写出并证明其对偶锥(dual cone)的具体形式。

2. （20分）证明：$f(x) = -(\Pi_{i=1}^n x_i)^{1/n}$, dom $f = \mathbb{R}_{++}^n$ 是凸函数。

3. （20分）给定$Q \in S^n$(实对称矩阵)，$\mu > 0$，$A_i \in \mathbb{R}^{m\times n}, i = 0, 1, \ldots, n$。对于矩阵$A \in \mathbb{R}^{m\times n}$，$\|A\|_2$为矩阵的二范数。令$A(x) = A_0 + \sum_{i=1}^n x_i A_i$。给出下列问题可以写成一个包含二次锥(second-order cone)和半定锥的凸优化问题的条件，并明确写出该凸优化的具体形式（需要具体写出二次锥和半定锥，严格给出等价关系）：

$$\min_x \quad \frac{\mu}{2} x^\top Q x + \|A(x)\|_2$$

4. （20分）写出下面问题的对偶问题及其最优性条件：

$$\begin{aligned} \min_{w,b,\xi} \quad & \|w\|_1 + C_1 \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i \cdot (a_i^\top w + b) \ge 1 - \xi_i, \forall i = 1, \ldots, n \\ & \xi_i \ge 0, \forall i = 1, \ldots, n \end{aligned}$$

5. （20分）给定$Q \in S^n$(实对称矩阵)，$b \in \mathbb{R}^m$和$a_i \in \mathbb{R}^n, i = 1, \ldots, m$，写出下面问题的对偶问题，以及该对偶问题的对偶问题：

$$\min_x \ x^\top Q x, \quad \text{s.t.} \quad (a_i^\top x)^2 = b_i, \ i = 1, \ldots, m.$$

21.5. Fourth

5. 去年，张师傅因为多旋圈面爆红，今年他来到了达摩院给扫地僧做面。某天，软件工程师小李跟张师傅吐槽工作。小李主要研究和设计算法用于调节各种产品的参数。这样的参数一般可以通过极小化 $\mathbb{R}^n$ 上的某个损失函数 f 求得。在小李最近的一个项目中，这个损失函数是另外一个课题组提供的；出于安全考虑和技术原因，该课题组难以向小李给出此函数的内部细节，而只能提供一个接口用于计算任意 $\mathbf{x} \in \mathbb{R}^n$ 处的函数值 $f(\mathbf{x})$。所以，小李必须仅基于函数值来极小化 f。而且，每次计算 f 的值都消耗不小的计算资源。好在该问题的维度 n 不是很高 (10 左右)。另外，提供函数的同事还告知小李不妨先假设 f 是光滑的。

这个问题让张师傅想起了自己收藏的一台古董收音机。要在这台收音机上收听一个节目，你需要小心地来回拧一个调频旋钮，同时注意收音效果，直到达到最佳。在这过程中，没有人确切地知道旋钮的角度和收音效果之间的定量关系是什么。张师傅和小李意识到，极小化 f 不过就是调节一台有多个旋钮的机器：想象 $\mathbf{x}$ 的每一个分量由一个旋钮控制，而 $f(\mathbf{x})$ 表示这台机器的某种性能，只要我们来回调整每个旋钮，同时监视 f 的值，应该就有希望找到最佳的 $\mathbf{x}$。受此启发，两人一起提出了极小化 f 的一个迭代算法，并命名为“自动前后调整算法” (Automated Forward/Backward Tuning，AFBT，算法 1)。

21.5. Fourth

"自动前后调整算法" (Automated Forward/Backward Tuning，AFBT，算法 1)。在第 k 次迭代中，AFBT 通过前后调整 $\mathbf{x}_k$ 的单个分量得到 $2n$ 个点 $\{\mathbf{x}_k \pm t_k\mathbf{e}^i : i = 1, \ldots, n\}$, 其中 t_k 为步长；然后，令 $\mathbf{y}_k$ 为这些点中函数值最小的一个，并检查 $\mathbf{y}_k$ 是否使 f 充分减小；若是，取 $\mathbf{x}_{k+1} = \mathbf{y}_k$，并将步长增倍；否则，令 $\mathbf{x}_{k+1} = \mathbf{x}_k$ 并将步长减半。在算法 1 中， $\mathbf{e}^i$ 表示 $\mathbb{R}^n$ 中的第 i 个坐标向量，它的第 i 个分量为 1, 其余皆为 0；$\mathbb{1}(\cdot)$ 为指示函数 —— 若 $f(\mathbf{x}_k) - f(\mathbf{y}_k)$ 至少为 t_k 之平方，则 $\mathbb{1}[f(\mathbf{x}_k) - f(\mathbf{y}_k) \geq t_k^2]$ 取值为 1，否则为 0。

1 自动前后调整算法 (AFBT)

输入 $\mathbf{x}_0 \in \mathbb{R}^n$, $t_0 > 0$。对 $k = 0, 1, 2, \ldots,$ 执行以下循环。

1: $\mathbf{y}_k := \operatorname{argmin}\{f(\mathbf{y}) : \mathbf{y} = \mathbf{x}_k \pm t_k\mathbf{e}^i,\ i = 1, \ldots, n\}$ # 计算损失函数。
2: $s_k := \mathbb{1}[f(\mathbf{x}_k) - f(\mathbf{y}_k) \geq t_k^2]$ # 是否充分下降？ 是：$s_k = 1$； 否：$s_k = 0$。
3: $\mathbf{x}_{k+1} := (1 - s_k)\mathbf{x}_k + s_k\mathbf{y}_k$ # 更新迭代点。
4: $t_{k+1} := 2^{2s_k - 1}t_k$ # 更新步长。$s_k = 1$：步长增倍； $s_k = 0$：步长减半。

现在，我们对损失函数 $f : \mathbb{R}^n \to \mathbb{R}$ 作出如下假设。

21.5. Fourth

假设 1. f 为凸函数，即对任何 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ 与 $\alpha \in [0, 1]$ 都有

$$f((1-\alpha)\mathbf{x} + \alpha\mathbf{y}) \;\le\; (1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{y}).$$

假设 2. f 在 $\mathbb{R}^n$ 上可微且 ∇f 在 $\mathbb{R}^n$ 上 L-Lipschitz 连续。

假设 3. f 的水平集有界，即对任意 $\lambda \in \mathbb{R}$，集合$\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \le \lambda\}$ 皆有界。

基于**假设 1** 与**假设 2**，可以证明

$$\langle \nabla f(\mathbf{x}), \, \mathbf{y} - \mathbf{x} \rangle \;\le\; f(\mathbf{y}) - f(\mathbf{x}) \;\le\; \langle \nabla f(\mathbf{x}), \, \mathbf{y} - \mathbf{x} \rangle + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2$$

对任何 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ 成立；**假设 1** 与**假设 3** 则保证f 在 $\mathbb{R}^n$ 上取到有限的最小值 f^*。凸函数的更多性质可参考任何一本凸分析教科书。

证明题(20分) 在**假设 1–3**下，对于 AFBT，证明

$$\lim_{k \to \infty} f(\mathbf{x}_k) \;=\; f^*.$$

R5 证明. 假设 $f(\mathbf{x}_k) \nrightarrow f^*$。记$\mathbf{g}_k = \nabla f(\mathbf{x}_k)$, 则 $\liminf \|\mathbf{g}_k\| > 0$ (取 $\mathbf{x}^*$ 使 $f(\mathbf{x}^*) = f^*$, 则由 f 之凸性知 $f(\mathbf{x}_k) - f^* \leq \langle \mathbf{g}_k, \mathbf{x}_k - \mathbf{x}^* \rangle$。因 $\{f(\mathbf{x}_k)\}$ 单调且 $f(\mathbf{x}_k) \nrightarrow f^*$，故 $\{\langle \mathbf{g}_k, \mathbf{x}_k - \mathbf{x}^* \rangle\}$ 有正下界。因 $\{f(\mathbf{x}_k)\}$ 不增且 f 水平集有界，故 $\{\mathbf{x}_k\}$ 有界，从而 $\liminf \|\mathbf{g}_k\| > 0$)。换言之，存在 $\epsilon > 0$ 使 $\|\mathbf{g}_k\| \geq \epsilon$ 对所有 $k \geq 0$ 成立。任给 $k \geq 0$, 可取 $i_k \in \{1, \ldots, n\}$ 满足

$$|\langle \mathbf{g}_k, \mathbf{e}^{i_k} \rangle| \ \geq \ \kappa \|\mathbf{g}_k\| \ \geq \ \kappa\epsilon, \tag{1}$$

其中 $\kappa = 1/\sqrt{n}$。故

$$f(\mathbf{y}_k) \ \leq \ \min\{f(\mathbf{x}_k \pm t_k \mathbf{e}^{i_k})\} \ \leq \ f(\mathbf{x}_k) - t_k |\langle \mathbf{g}_k, \mathbf{e}^{i_k} \rangle| + \frac{L t_k^2}{2} \ \leq \ f(\mathbf{x}_k) - \kappa\epsilon t_k + \frac{L t_k^2}{2}. \tag{2}$$

所以，只要

$$t_k \ \leq \ \frac{2\kappa\epsilon}{L+2}, \tag{3}$$

我们就有 $f(\mathbf{y}_k) \leq f(\mathbf{x}_k) - t_k^2$, 从而 $s_k = 1$, 进而有 $t_{k+1} = 2t_k$。由此，易见

$$t_k \ \geq \ \underline{t} \ \equiv \ \min\left\{t_0, \ \frac{\kappa\epsilon}{L+2}\right\} \ > \ 0 \tag{4}$$

对所有 $k \geq 0$ 成立。故存在无穷个 k 使得 $s_k = 1$ (否则 $t_k \to 0$); 对每一个这样的 k, $f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq \underline{t}^2$。这与 f 下方有界相矛盾。故所论成立。

21.6. References

[1] https://web.stanford.edu/class/ee364a/homework.html

[2] http://bicmr.pku.edu.cn/~wenzw/opt-2020-fall.html

[3] https://www.mathworks.com/help/matlab/math/voronoi-diagrams.html

[4] http://bicmr.pku.edu.cn/~wenzw/opt2015/mid-2019.pdf