第 1、2 章

1. 集合（线性空间、闭集合、开集合、仿射集、凸集、锥、凸锥、半空间、超平面、球、椭圆、范数球、范数锥、多面体、半正定锥、正定锥等）
2. 线性映射、线性无关、线性相关、基、维数
3. 范数（定义、性质和应用）、常见范数、对偶范数
4. 保凸运算（交集、仿射函数、线性分式及透视函数）
5. 分割超平面定理、支撑超平面定理
6. 正常锥与广义不等式（最小元与极小元）、对偶锥

第 3 章

1. 概念：凸函数的定义、扩展值函数、凸函数的一阶条件定义、凸函数的二阶条件定义、下水平集、上镜图、Jensen's 不等式
2. 凸函数的判断（定义、性质、上镜图、保凸运算）
3. 保凸运算（非负加权求和、带有仿射映射的组合、逐点最大值、逐点上确界、标量函数组合、向量函数组合、最小化、透视函数）
4. 共轭函数
5. 拟凸函数、对数凹（凸）函数

第 4 章

1. 概念：目标函数、优化变量、不等式约束、等式约束、可行解、最优值、最优点（平凡定义）、局部最优点、全局最优点、优化问题的标准形式
2. 凸优化问题的标准形式、最优条件、可微目标函数优化问题的最优条件（简

单约束问题）

3. 优化问题的等价变换：变量变换、函数变换、松弛变量、消除等式约束、引入等式约束、上镜图、取最小值
4. 拟凸函数与拟凸优化问题
5. 凸优化的典型问题：线性规划、二次规划、几何规划

第 5 章

1. 概念：Lagrangian、Lagrange 乘子、Lagrange 对偶函数、最优值的下界、Lagrange 对偶问题、对偶约束条件
2. 对偶转化：引入新变量和等式约束、目标函数变换、隐藏约束
3. 对偶性：弱对偶性、强对偶性，、Slater 约束准则（几何解释）、互补松弛条件
4. KKT 条件及其运用

第 6 章

1. 逼近问题、随机估计问题
2. 最速下降方法、牛顿法

凸集的保凸性

2. 设 $C \subset \mathbb{R}^p$ 是一个凸集, p 是正整数. 证明下列集合 S 是 $\mathbb{R}^n$ 中的凸集:

$$S = \{\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} = \boldsymbol{A}\rho, \rho \in C\}$$

其中 $\boldsymbol{A}$ 是给定的 $n \times p$ 实矩阵.

证明 证对任意两点 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in S$ 及每个数 $\lambda \in [0,1]$, 根据集合 S 的定义, 存在 $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2 \in C$, 使 $\boldsymbol{x}^{(1)} = \boldsymbol{A}\boldsymbol{\rho}_1, \boldsymbol{x}^{(2)} = \boldsymbol{A}\boldsymbol{\rho}_2$, 因此必有 $\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)} = \lambda\boldsymbol{A}\boldsymbol{\rho}_1+(1-\lambda)\boldsymbol{A}\boldsymbol{\rho}_2 = \boldsymbol{A}\left[\lambda\boldsymbol{\rho}_1 + (1-\lambda)\boldsymbol{\rho}_2\right]$. 由于 C 是凸集, 必有 $\lambda\rho_1 + (1-\lambda)\rho_2 \in C$, 因此 $\lambda x^{(1)} + (1-\lambda)x^{(2)} \in S$, 故 S 是凸集.

凸函数的证明（上镜图）

3.3 *Inverse of an increasing convex function.* Suppose $f : \mathbf{R} \to \mathbf{R}$ is increasing and convex on its domain (a, b). Let g denote its inverse, *i.e.*, the function with domain $(f(a), f(b))$ and $g(f(x)) = x$ for $a < x < b$. What can you say about convexity or concavity of g?

Solution. g is concave. Its hypograph is

$$\begin{aligned}
\textbf{hypo}\, g &= \{(y,t) \mid t \le g(y)\} \\
&= \{(y,t) \mid f(t) \le f(g(y))\} \quad \text{(because } f \text{ is increasing)} \\
&= \{(y,t) \mid f(t) \le y)\} \\
&= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \textbf{epi}\, f.
\end{aligned}$$

For differentiable g, f, we can also prove the result as follows. Differentiate $g(f(x)) = x$ once to get

$$g'(f(x)) = 1/f'(x).$$

so g is increasing. Differentiate again to get

$$g''(f(x)) = -\frac{f''(x)}{f'(x)^3},$$

so g is concave.

凸函数的一阶条件定义（性质）

3.1 Suppose $f : \mathbf{R} \to \mathbf{R}$ is convex, and $a,\ b \in \mathbf{dom}\, f$ with $a < b$.

(a) Show that

$$f(x) \le \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b)$$

for all $x \in [a, b]$.

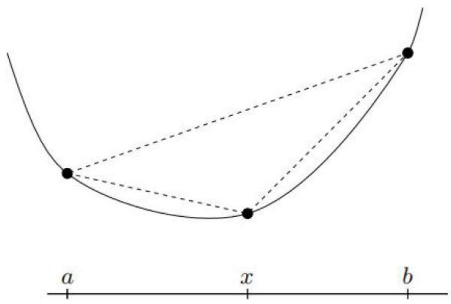
Solution. This is Jensen's inequality with $\lambda = (b-x)/(b-a)$.

(b) Show that

$$\frac{f(x) - f(a)}{x - a} \le \frac{f(b) - f(a)}{b - a} \le \frac{f(b) - f(x)}{b - x}$$

for all $x \in (a, b)$. Draw a sketch that illustrates this inequality.

Solution. We obtain the first inequality by subtracting $f(a)$ from both sides of the inequality in (a). The second inequality follows from subtracting $f(b)$. Geometrically, the inequalities mean that the slope of the line segment between $(a, f(a))$ and $(b, f(b))$ is larger than the slope of the segment between $(a, f(a))$ and $(x, f(x))$, and smaller than the slope of the segment between $(x, f(x))$ and $(b, f(b))$.



(c) Suppose f is differentiable. Use the result in (b) to show that

$$f'(a) \le \frac{f(b) - f(a)}{b - a} \le f'(b).$$

Note that these inequalities also follow from (3.2):

$$f(b) \ge f(a) + f'(a)(b - a), \qquad f(a) \ge f(b) + f'(b)(a - b).$$

Solution. This follows from (b) by taking the limit for $x \to a$ on both sides of the first inequality, and by taking the limit for $x \to b$ on both sides of the second inequality.

(d) Suppose f is twice differentiable. Use the result in (c) to show that $f''(a) \ge 0$ and $f''(b) \ge 0$.

Solution. From part (c),

$$\frac{f'(b) - f'(a)}{b - a} \ge 0,$$

and taking the limit for $b \to a$ shows that $f''(a) \ge 0$.

3.47 Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is differentiable, $\mathbf{dom}\, f$ is convex, and $f(x) > 0$ for all $x \in \mathbf{dom}\, f$. Show that f is log-concave if and only if for all $x, y \in \mathbf{dom}\, f$,

$$\frac{f(y)}{f(x)} \le \exp\left(\frac{\nabla f(x)^T (y - x)}{f(x)}\right).$$

Solution. This is the basic inequality

$$h(y) \ge h(x) + \nabla h(x)^T (y - x)$$

applied to the convex function $h(x) = -\log f(x)$, combined with $\nabla h(x) = (1/f(x)) \nabla f(x)$.

3.5 [RV73, 第 22 页] 凸函数的滑动平均。假设 $f : \mathbf{R} \to \mathbf{R}$ 是凸函数，且 $\mathbf{R}_+ \subseteq \mathbf{dom}\, f$。证明其滑动平均 F 是凸的，其中 F 定义为

$$F(x) = \frac{1}{x} \int_0^x f(t)\, dt, \quad \mathbf{dom}\, F = \mathbf{R}_{++},$$

并假设 f 是可微的。

证明 F 是可微的，其导数为

$$\begin{aligned}
F'(x) &= -\frac{1}{x^2} \int_0^x f(t)\, dt + \frac{f(x)}{x}, \\
F''(x) &= \frac{2}{x^3} \int_0^x f(t)\, dt - \frac{2f(x)}{x^2} + \frac{f'(x)}{x} \\
&= \frac{2}{x^3} \int_0^x \left(f(t) - f(x) - f'(x)(t-x) \right) dt.
\end{aligned}$$

由以下事实可得 F 的凸性：

$$f(t) \geq f(x) + f'(x)(t-x)$$

对所有 $x, t \in \mathbf{dom}\, f$ 都成立，这意味着 $F''(x) \geq 0$。

凸优化问题的最优条件

4.3 Prove that $x^\star = (1, 1/2, -1)$ is optimal for the optimization problem

$$\begin{array}{ll} \text{minimize} & (1/2)x^TPx + q^Tx + r \\ \text{subject to} & -1 \leq x_i \leq 1, \quad i = 1,2,3 \end{array}$$

where

$$P = \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix}, \quad q = \begin{bmatrix} -22.0 \\ -14.5 \\ 13.0 \end{bmatrix}, \quad r = 1.$$

Solution. We verify that $x^\star$ satisfies the optimality condition (4.21). The gradient of the objective function at $x^\star$ is

$$\nabla f_0(x^\star) = (-1, 0, 2)$$

Therefore the optimality condition is that

$$\nabla f_0(x^\star)^T (y - x) = -1(y_1 - 1) + 2(y_2 + 1) \geq 0$$

for all y satisfying $-1 \leq y_i \leq 1$, which is clearly true.

线性规划

4.9 Square LP. Consider the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

with A square and nonsingular. Show that the optimal value is given by

$$p^\star = \begin{cases} c^T A^{-1} b & A^{-T} c \preceq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Solution. Make a change of variables $y = Ax$. The problem is equivalent to

$$\begin{array}{ll} \text{minimize} & c^T A^{-1} y \\ \text{subject to} & y \preceq b. \end{array}$$

If $A^{-T} c \preceq 0$, the optimal solution is $y = b$, with $p^\star = c^T A^{-1} b$. Otherwise, the LP is unbounded below.

对偶问题

5.3 Problems with one inequality constraint. Express the dual problem of

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & f(x) \le 0, \end{array}$$

with $c \neq 0$, in terms of the conjugate f^*. Explain why the problem you give is convex. We do not assume f is convex.

Solution.

For $\lambda = 0, g(\lambda) = \inf c^T x = -\infty$. For $\lambda > 0$,

$$\begin{aligned} g(\lambda) &= \inf\left(c^T x + \lambda f(x)\right) \\ &= \lambda \inf\left((c/\lambda)^T x + \lambda f(x)\right) \\ &= -\lambda f_1^*(-c/\lambda), \end{aligned}$$

i.e., for $\lambda \ge 0, -g$ is the perspective of f_1^*, evaluated at $-c/\lambda$. The dual problem is

$$\begin{array}{ll} \text{minimize} & -\lambda f_1^*(-c/\lambda) \\ \text{subject to} & \lambda \ge 0. \end{array}$$

5.15 Consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \end{array}$$

where the functions $f_i : \mathrm{R}^n \to \mathrm{R}$ are differentiable and convex. Let $h_1, \ldots, h_m : \mathrm{R} \to \mathrm{R}$ be increasing differentiable convex functions. Show that

$$\phi(x) = f_0(x) + \sum_{i=1}^{m} h_i\left(f_i(x)\right)$$

is convex. Suppose $\bar{x}$ minimizes ϕ. Show how to find from $\bar{x}$ a feasible point for the dual of (5.109). Find the corresponding lower bound on the optimal value of (5.109).

Solution.

$\bar{x}$ satisfies

$$0 = \nabla f_0(\bar{x}) + \sum_{i=1}^{m} \left(h_i'\left(f_i(\bar{x})\right)\right) \nabla f_i(\bar{x})) = \nabla f_0(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\bar{x}))$$

where $\lambda_i = h_i'\left(f_i(\bar{x})\right).\lambda$ is dual feasible: $\lambda_i \geq 0$, since h_i is increasing, and

$$\begin{aligned} g(\lambda) &= f_0(\bar{x}) + \sum_{i=1}^{m} \lambda_i f_i(\bar{x}) \\ &= f_0(\bar{x}) + \sum_{i=1}^{m} h_i'\left(f_i(\bar{x})\right) f_i(\bar{x}) \end{aligned}$$

5.16 An exact penalty method for inequality constraints. Consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m \end{array}$$

where the functions $f_i : \mathbf{R}^n \to \mathbf{R}$ are differentiable and convex. In an exact penalty method, we solve the auxiliary problem

$$\text{minimize} \quad \phi(x) = f_0(x) + \alpha \max_{i=1}, \ldots, m \max\{0, f_i(x)\}$$

where $\alpha > 0$ is a parameter. The second term in ϕ penalizes deviations of x from feasibility. The method is called an exact penalty method if for sufficiently large α, solutions of the auxiliary problem (5.111) also solve the original problem (5.110).

(a) Show that ϕ is convex.

(b) The auxiliary problem can be expressed as

$$\begin{array}{ll} \text{minimize} & f_0(x) + \alpha y \\ \text{subject to} & f_i(x) \le y, \quad i = 1, \ldots, m \\ & 0 \le y \end{array}$$

where the variables are x and $y \in \mathbf{R}$. Find the Lagrange dual of this problem, and express it in terms of the Lagrange dual function g of (5.110).

(c) Use the result in (b) to prove the following property. Suppose $\lambda^\star$ is an optimal solution of the Lagrange dual of (5.110), and that strong duality holds. If $\alpha > \mathbf{1}^T \lambda^\star$, then any solution of the auxiliary problem (5.111) is also an optimal solution of (5.110).

Solution.

(a) The first term is convex. The second term is convex since it can be expressed as

$$\max\{f_1(x), \ldots, f_m(x), 0\}$$

i.e., the pointwise maximum of a number of convex functions.

(b) The Lagrangian is

$$L(x, y, \lambda, \mu) = f_0(x) + \alpha y + \sum_{i=1}^m \lambda_i (f_i(x) - y) - \mu y$$

The dual function is

$$\begin{aligned} \inf_{x,y} L(x, y, \lambda, \mu) &= \inf_{x,y} f_0(x) + \alpha y + \sum_{i=1}^m \lambda_i (f_i(x) - y) - \mu y \\ &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \right) + \inf_y \left(\alpha - \sum_{i=1}^m \lambda_i - \mu \right) y \\ &= \begin{cases} g(\lambda) & \mathbf{1}^T \lambda + \mu = \alpha \\ -\infty & \text{otherwise,} \end{cases} \end{aligned}$$

and the dual problem is

$$\begin{array}{ll} \text{maximize} & g(\lambda) \\ \text{subject to} & \mathbf{1}^T \lambda + \mu = \alpha \\ & \lambda \succeq 0, \quad \mu \ge 0 \end{array}$$

or, equivalently,

$$\begin{array}{ll} \text{maximize} & g(\lambda) \\ \text{subject to} & \mathbf{1}^T \lambda \le \alpha \\ & \lambda \succeq 0. \end{array}$$

(c) If $\mathbf{1}^T \lambda^\star < \alpha$, then $\lambda^\star$ is also optimal for the dual problem derived in part (b). By complementary slackness $y = 0$ in any optimal solution of the primal problem, so the optimal x satisfies $f_i(x) \le 0, i = 1, \ldots, m$, i.e., it is feasible in the original problem, and therefore also optimal.

KKT

5.29 The problem

$$\begin{array}{ll} \text{minimize} & -3x_1^2 + x_2^2 + 2x_3^2 + 2(x_1 + x_2 + x_3) \\ \text{subject to} & x_1^2 + x_2^2 + x_3^2 = 1 \end{array}$$

is a special case of (5.32), so strong duality holds even though the problem is not convex. Derive the KKT conditions. Find all solutions x, ν that satisfy the KKT conditions. Which pair corresponds to the optimum?

Solution.

(a) The KKT conditions are

$$x_1^2 + x_2^2 + x_3^2 = 1, \quad (-3 + \nu)x_1 + 1 = 0, \quad (1 + \nu)x_2 + 1 = 0, \quad (2 + \nu)x_3 + 1 = 0$$

(b) A first observation is that the KKT conditions imply $\nu \neq 2, \nu \neq -1, \nu \neq 3$. We can therefore eliminate x and reduce the KKT conditions to a nonlinear equation in ν :

$$\frac{1}{(-3+\nu)^2} + \frac{1}{(1+\nu)^2} + \frac{1}{(2+\nu)^2} = 1$$

The lefthand side is plotted in the figure 5.

There are four solutions:

$$\nu = -3.15, \quad \nu = 0.22, \quad \nu = 1.89, \quad \nu = 4.04$$

corresponding to

$$\begin{array}{ll} x = (0.16, 0.47, -0.87), & x = (0.36, -0.82, 0.45), \\ x = (0.90, -0.35, 0.26), & x = (-0.97, -0.20, 0.17) \end{array}$$

(c) $\nu^\star$ is the largest of the four values: $\nu^\star = 4.0352$. This can be seen several ways. The simplest way is to compare the objective values of the four solutions x, which are

$$f_0(x) = 1.17, \quad f_0(x) = 0.67, \quad f_0(x) = -0.56, \quad f_0(x) = -4.70$$

5.30 Derive the KKT conditions for the problem

$$\begin{array}{ll} \text{minimize} & \operatorname{tr} X - \log\det X \\ \text{subject to} & Xs = y, \end{array}$$

with variable $X \in \mathbf{S}^n$ and domain $\mathbf{S}^n_{++}$.$y \in \mathbf{R}^n$ and $s \in \mathbf{R}^n$ are given, with $s^T y = 1$. Verify that the optimal solution is given by

$$X^\star = I + yy^T - \frac{1}{s^T s} ss^T$$

Solution.
We introduce a Lagrange multiplier $z \in \mathbf{R}^n$ for the equality constraint. The KKT optimality conditions are:

$$X \succ 0, \quad Xs = y, \quad X^{-1} = I + \frac{1}{2}\left(zs^T + sz^T\right)$$

We first determine z from the condition $Xs = y$. Multiplying the gradient equation on the right with y gives

$$s = X^{-1}y = y + \frac{1}{2}\left(z + \left(z^T y\right) s\right)$$

By taking the inner product with y on both sides and simplifying, we get $z^T y = 1 - y^T y$. Substituting in (5.30.B) we get

$$z = -2y + \left(1 + y^T y\right) s$$

and substituting this expression for z in (5.30.A) gives

$$\begin{aligned} X^{-1} &= I + \frac{1}{2}\left(-2ys^T - 2sy^T + 2\left(1 + y^T y\right) ss^T\right) \\ &= I + \left(1 + y^T y\right) ss^T - ys^T - sy^T \end{aligned}$$

Finally, we verify that this is the inverse of the matrix $X^\star$ given above:

$$\begin{aligned} &\left(I + \left(1 + y^T y\right) ss^T - ys^T - sy^T\right) X^\star \\ &= \left(I + yy^T - \left(1/s^T s\right) ss^T\right) + \left(1 + y^T y\right)\left(ss^T + sy^T - ss^T\right) \\ &\quad - \left(ys^T + yy^T - ys^T\right) - \left(sy^T + \left(y^T y\right) sy^T - \left(1/s^T s\right) ss^T\right) \\ &= I. \end{aligned}$$

To complete the solution, we prove that $X^\star \succ 0$. An easy way to see this is to note that

$$X^\star = I + yy^T - \frac{ss^T}{s^T s} = \left(I + \frac{ys^T}{\|s\|_2} - \frac{ss^T}{s^T s}\right)\left(I + \frac{ys^T}{\|s\|_2} - \frac{ss^T}{s^T s}\right)^T$$