总成绩 (100%) = 考勤 (10%) + 习题 (30%) + 测验 (60%)，习题每个点至少完成一半

课程概况：

- 绪论 2h
- 凸集和凸函数 3h
- 凸优化问题 6h
- 拉格朗日乘子 5h
- 凸优化应用 6h
- 无约束凸优化问题求解 5h
- 有约束凸优化问题求解 4h
- 课程测试 2h

1 绪论

Remark 1.1. f 是凸函数，$\nabla f_x = 0 \iff$ 点 x 是最小值点

2 凸集和凸函数

Remark 2.1. 凸集：$\forall x, y \in C, \theta \in [0,1] \Longrightarrow \theta x + (1-\theta)y \in C$。空集、点、线段都是凸集。

Remark 2.2. 凸集的例子：

- 超平面：$\{x : a^t x = b\}$，$a \in \mathbb{R}^n - \{0\}$ 是法向量
- 半平面：$\{x : a^t x \leq b\}$，$a \in \mathbb{R}^n - \{0\}$
- 欧几里得球：$B(x_c, r) = \{x : \|x - x_c\| < r\} = \{x_c + ru : \|u\| < 1\}$
- 椭球：$\{x : (x - x_c)^t P^{-1}(x - x_c) < 1\}, P \in S^n_{++}$ 或 $\{x_c + Au : \|u\| < 1\}, A \in S^n_{++}$
- 多面体：$\{x : Ax \preceq b, Cx = d\}, A \in \mathbb{R}^{m\times n}, C \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^m, d \in \mathbb{R}^p$，多面体是半空间和超平面的交集，任意个数的凸集的交集是凸集。

Remark 2.3. 凸集 S：是 S 中所有点的凸组合的最小集合

- 如果 C 是一个凸集，则 $aC + b = \{ax + b : x \in C\}, a \in \mathbb{R}, b \in \mathbb{R}^n$ 也是一个凸集
- 对于仿射函数 $f(x) = Ax + b, A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$

  - 凸集在 f 下的像是凸集

$$S \subset \mathbb{R}^n \text{ convex } \Longrightarrow f(S) = \{f(x) : x \in S\} \subset \mathbb{R}^m \text{ convex}$$

  - 凸集在 f 下的逆像是凸集

$$C \subset \mathbb{R}^m \text{ convex } \Longrightarrow f^{-1}(C) = \{x : f(x) \in C\} \subset \mathbb{R}^n \text{ convex}$$

- 两个凸集可以用一个超平面分离（证明困难）
- 支撑超平面：$\{x : a^t x = a^t x_0\}, a \in \mathbb{R}^n - \{0\}, a^t x \le a^t x_0, \forall x \in C$，其中 C 是一个凸集，x_0 是凸集上一边界点
- 支撑超平面定理：如果 C 是凸的，则在 C 的每个边界点上都存在一个支撑超平面

Remark 2.4. 凸函数：$f : \mathbb{R}^n \to \mathbb{R}$ 的定义域 $\operatorname{dom}(f)$ 是一个凸集，满足 $\forall x, y \in \operatorname{dom}(f), \theta \in [0, 1]$

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

f 是一个凸函数，则 f 在定义域的任何内点都是连续的，并且 f 是局部有界的：$\exists B(x, r) \subset \operatorname{dom}(f)$

$$z = \theta x + (1-\theta)y \Longrightarrow \frac{f(z) - f(x)}{1-\theta} \le \frac{f(y) - f(z)}{\theta} \Longrightarrow \frac{f(z) - f(x)}{\|z - x\|} \le \frac{f(y) - f(z)}{\|y - z\|}$$

Remark 2.5. 一些凸函数的例子：

- 仿射函数：$a^t x + b, \forall a \in \mathbb{R}^n, b \in \mathbb{R}$
- 幂函数 x^α on $\mathbb{R}_{++} = (0, \infty), \alpha \ge 1$ or $\alpha \le 0$
- l^p 范数：$\|x\|_p = \sum_{i=1}^n (|x_i|^p)^{\frac{1}{p}}$ on $\mathbb{R}^n$ for $p \ge 1$

Remark 2.6. 凸函数的性质

- f is convex $\Longrightarrow \alpha f$ for $\alpha > 0$ is convex.
- $f_1, \ldots, f_m$ are convex $\Longrightarrow f_1 + \cdots + f_m$ is convex.
- Composition with affine function: f is convex $\Longrightarrow f(Ax + b)$ is convex.
- $f_1, \ldots, f_m$ are convex $\Longrightarrow f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex.
  - 分段线性函数 (piecewise-linear function): $f(x) = \max_{1 \le i \le m} \{a_i^t x + b_i\}$

Remark 2.7. 严格凸函数 strictly convex function，$f : \mathbb{R}^n \to \mathbb{R}$

- $\mathrm{dom}(f)$ is convex set
- $\forall x \neq y \in \mathrm{dom}(f), \theta \in (0,1)$

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

l^p-norm and l^∞-norm are not strictly convex.

Remark 2.8. 强凸函数 strongly convex function，$f : \mathbb{R}^n \to \mathbb{R}$

- $\mathrm{dom}(f)$ is convex set
- $\exists m > 0$, satisfy $f(x) - \frac{m}{2}\|x\|^2$ is convex.
- 判定方法：$\nabla^2 f \succ 0$
- 谱分解：$A = P^t Q P$，其中 Q 的对角线是 A 的特征值
- $\nabla^2 f - mI \succeq 0 \Longrightarrow u^t(\nabla^2 f - mI)u \geq 0 \Longrightarrow u^t \nabla^2 f u \geq m\|u\|^2 \Longrightarrow u^t P^t Q P u \geq m\|u\|^2 \Longrightarrow \forall \lambda \geq m, \lambda$ 是 $\nabla^2 f(x)$ 的特征值
- $f(x)$ is convex $\Longrightarrow f(x) + \frac{m}{2}\|x\|^2$ is strictly convex.

strongly convex $\Longrightarrow$ strictly convex $\Longrightarrow$ convex

Remark 2.9. 凸函数判定：

- First-order condition: **differentiable** f with **convex domain** is convex iff $\forall x, y \in \mathrm{dom}(f)$

$$f(y) \geq f(x) + \nabla f(x)^t (y - x)$$

  - *Proof.*
    * f is convex $\Longleftrightarrow \frac{f(z)-f(x)}{z-x} \leq \frac{f(y)-f(z)}{y-z} \Longrightarrow$ with $z \to x$, $\frac{f(y)-f(x)}{y-x} \geq f'(x) \Longrightarrow f(y) \geq f(x) + f'(x)(y-x)$
    * $f(y) \geq f(x) + f'(x)(y-x) \Longrightarrow \frac{f(y)-f(z)}{y-z} \geq f'(z) \geq \frac{f(x)-f(z)}{x-z} \Longrightarrow$ by $z = \theta x + (1-\theta)y, f$ is convex

    □

  - $f(x^*) = 0 \Longleftrightarrow x^*$ is global minimum of f.
  - f is strictly convex $\Longleftrightarrow \forall x \neq y \in \mathrm{dom}(f), f(y) > f(x) + \nabla f(x)^t (y-x)$

- Second-order conditions: for **twice differentiable** f with **convex domain** is convex iff $x \in \mathrm{dom}(f)$

$$\nabla^2 f(x) \succeq 0$$

- if $\nabla^2 f(x) \succ 0$ for $\forall x \in \mathrm{dom}(f) \Longrightarrow f$ is strictly convex.
- $\exists m > 0$, satisfy $\nabla^2 f(x) \succeq mI$ for $\forall x \in \mathrm{dom}(f) \Longleftrightarrow f$ is strongly convex.

- Restriction of a convex function to a line:
  - $f : \mathbb{R}^n \to \mathbb{R}$ is convex iff the function $g : \mathbb{R} \to \mathbb{R}$

$$g(t) = f(x + tv), \quad \mathrm{dom}(g) = \{t : x + tv \in \mathrm{dom}(f)\}$$

    is convex for any $x \in \mathrm{dom}(f)$ and $v \in \mathbb{R}^n$
  - *Proof.* $g(t) = f(x + t(y - x))$ is convex $\Longleftrightarrow g(\theta) \leq \theta g(0) + (1-\theta)g(1) \Longleftrightarrow f(\theta x + (1-\theta y)) \leq \theta f(x) + (1-\theta)f(y) \Longleftrightarrow f$ is convex. □

Remark 2.10. $X \in \mathbb{S}^n_{++} \Longrightarrow X = P^t Q P = P^t \operatorname{diag}(q_1, \ldots, q_n) P \Longrightarrow X^\alpha \triangleq P^t \operatorname{diag}(q_1^\alpha, \ldots, q_n^\alpha) P$, satisfy $X^\alpha X^\beta = X^{\alpha+\beta}$, $X^0 = I$.

Remark 2.11.

- sublevel set(下水平集) of $f : \mathbb{R}^n \to \mathbb{R}$

$$C_\alpha = \{x \; in \, \mathrm{dom}(f) : f(x) \leq \alpha\}$$

  sublevel set of convex functions are convex.

- Epigraph(上境图) set of $f : \mathbb{R}^n \to \mathbb{R}$

$$\mathrm{epi}(f) = \big\{(x, t) \in \mathbb{R}^{n+1} : x \in \mathrm{dom}(f), t \geq f(x)\big\}$$

  f is convex iff epi is convex set.

3 凸优化问题

Remark 3.1. Optimization problem:

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & f_i(x) \leq 0, \quad 1 \leq i \leq m \\ & h_i(x) = 0, \quad 1 \leq i \leq p \end{cases}$$

- feasible set $X \subset \mathcal{D}$

$$\mathcal{D} = \mathrm{dom}(f) \cap \left(\cap_{i=1}^m \mathrm{dom}\,(f_i)\right) \cap \left(\cap_{i=1}^p \mathrm{dom}\,(h_i)\right)$$

- optimal value: $p^* = \inf\{f(x) : x \text{ is feasible}\}$

- a feasible x is an optimal solution(minimizer) if $f(x) = p^*$

Remark 3.2. Convex optimization problem(COP):

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & f_i(x) \le 0, \quad 1 \le i \le m \\ & Ax = b \end{cases}$$

- objective function f is convex
- inequality constraints $f_1, \ldots, f_m$ are convex
- equality constraints are affine: $A \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^p$
- the feasible set X of COP is convex
  - $X = \mathrm{dom}(f) \cap (\cap_{i=1}^m X_i) \cap \text{hyperplanes}$
    - $X_i = \{x \in \mathrm{dom}(f_i) : f_i(x) \le 0\}$
  - so a COP is actually an unconstrained COP defined on a convex set.
- any local minimum of a COP is globally optimal.
  - x^* is a local minimum: a solution of the COP in $B(x^*, r) \cap X$
  - $\forall y \in X$, take $\theta \to 1$, satisfy $z = \theta x^* + (1-\theta)y \in B(x^*, r)$, by convexity, $\theta f(x^*) + (1-\theta)f(y) \ge f(z) \ge f(x^*)$, thus $f(y) \ge f(x^*)$.
- the set of optimal solutions is convex.
- For **differentiable** f, $x \in X$ is optimal iff

$$\nabla f(x)^t (y - x) \ge 0, \quad \forall y \in X$$

  - if x is optimal, let $g(\theta) = f(x + \theta(y - x))$

$$0 \le \lim_{\theta \downarrow 0} \frac{f(x + \theta(y-x)) - f(x)}{\theta} = g'(0) = \nabla f(x)^t (y - x)$$

  - conversely, by convexity(First order condition)

$$\begin{cases} f(y) \ge f(x) + \nabla f(x)^t (y - x) \\ \nabla f(x)^t (y - x) \ge 0 \end{cases} \Longrightarrow f(y) \ge f(x) \Longrightarrow x \text{ is optimal}$$

Important, for COP:

$$x \text{ optimal} \Longleftrightarrow \nabla f(x)^t (y - x) \ge 0, \forall y \in X$$

Remark 3.3. Important examples:

- Linear program(LP):

$$\begin{cases} \text{minimize} & c^t x \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{cases}$$

  - convex problem with affine object over a polyhedron
  - standard from

$$\begin{cases} \text{minimize} & c^t x \\ \text{subject to} & x \succeq 0 \\ & Ax = b \end{cases}$$

- Quadratic program(QP):

$$\begin{cases} \text{minimize} & \frac{1}{2}x^t Px + q^t x + r \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{cases}$$

  $P \in \mathbb{S}_+^n$, convex problem with quadratic object over a polyhedron.

- Quadratically constrained quadratic program(QCQP)

$$\begin{cases} \text{minimize} & \frac{1}{2}x^t Px + q^t x + r \\ \text{subject to} & \frac{1}{2}x^t P_i x + q_i^t x + r \preceq 0, \quad 1 \le i \le m \\ & Ax = b \end{cases}$$

  - $P, P_i \in \mathbb{S}_+^n$, objective and constraints are convex quadratic.
  - if $P_1, \ldots, P_m \in \mathbb{S}_{++}^n$, feasible region is intersection of m ellipsoids andan affine set.

Remark 3.4. Unconstrained COP, with differentiable f

$$\text{minimize}\, f(x)$$

- $x \in \operatorname{dom}(f)$ (open set!) is optimal iff $\nabla f(x) = 0$
  - if x is optimal, $\nabla f(x)^t(y - x) \ge 0$ for any feasible y, take $y = x - \lambda \nabla f(x)$ for sufficient small $\lambda > 0$, thus $\nabla f(x) = 0$
  - conversely, $\nabla f(x)^t(y - x) = 0$

- Intuitive interpretation: x is optimal, then $\langle \nabla f(x), y-x\rangle \geq 0$. if $\nabla f(x) \neq 0$, $\exists y$ satisfy $\langle \nabla f(x), y-x\rangle < 0$, so $\nabla f(x) = 0$.

Remark 3.5. Equality constrained COP, with differentiable f, $A \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^p$

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{cases}$$

- $x \in \operatorname{dom}(f)$ is optimal iff:

$$Ax = b, \text{ and there exists } v \in \mathbb{R}^p, \text{ s.t. } \nabla f(x) = A^t v, \nabla f(x) \in \mathcal{R}(A^t)$$

Proof. *x optimal* $\Longleftrightarrow Ax = b, \nabla f(x) = A^t v$.

- "$\Longleftarrow$": $\forall y \in X$, $Ay = b \Longrightarrow (\nabla f(x))^t(y-x) = v^t A(y-x) = v^t(b-b) = 0 \Longrightarrow x$ is optimal.
- $\mathcal{N}(A) = \mathcal{R}(A^t)^\perp, \mathcal{N}(A)^\perp = \mathcal{R}(A^\perp)$
- "$\Longrightarrow$": $\forall u \in \mathcal{N}(A)$, $x + \theta u \in \operatorname{dom}(f)$ for $\theta \to 0$. make $y = x + \theta u$, $\nabla f(x)(y-x) \geq 0 \Longrightarrow \theta\langle \nabla f(x), u\rangle \geq 0$ satisfy for all $\theta \to 0$, so $\langle \nabla f(x), u\rangle = 0$. As a result, $\nabla f(x)^t \in \mathcal{N}(A)^\perp$, then $\exists v \in \mathbb{R}^p$ s.t. $\nabla f(x) = A^t v$.

□

Remark 3.6. Equality constrained QP: $P \in \mathbb{S}^n_+, q \in \mathbb{R}^n, r \in \mathbb{R}, A \in \mathbb{R}^{p\times n}$ with $\operatorname{rank}(A) = p$, $b \in \mathbb{R}^p$

$$\begin{cases} \text{minimize} & \frac{1}{2}x^t P x + q^t x + r \\ \text{subject to} & Ax = b \end{cases}$$

- x^* is optimal $\Longleftrightarrow \exists v^* \in \mathbb{R}^p$ s.t.

$$\begin{bmatrix} P & A^t \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ v^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

- coefficient matrix is called KKT matrix.
- KKT matrix is nonsingular $\Longleftrightarrow$ "$Ax = 0, x \neq 0 \Longrightarrow x^t P x > 0$"

Remark 3.7.

$$\begin{cases} \text{minimize}_x & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad 1 \leq i \leq m \end{cases} \Longleftrightarrow \begin{cases} \text{minimize}_{(x,y)} & f(x) \\ \text{subject to} & g_i(x) + y_i^2 = 0, \quad 1 \leq i \leq m \end{cases}$$

- $L(x, y, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i(g_i(x) + y_i^2), \quad \partial_x L = \partial_y L = 0$
- $\begin{cases} \nabla f(x) + \sum \lambda_i \nabla g_i(x) = 0 \\ \lambda_i y_i = 0 \end{cases} \Longrightarrow \begin{cases} \lambda_i = 0 \\ y_i = 0 \end{cases} \Longrightarrow \lambda_i g_i(x) = 0$

Remark 3.8. Inequality constrained COP, with differentiable $f, f_1, \ldots, f_m$

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & f_i(x) \leq 0, \quad 1 \leq i \leq m \end{cases}$$

- sufficient condition: for a feasible x, if exists $\lambda_i \geq 0$ for $i \in [1, m]$ and $\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) = 0$, $\lambda_1 f_1(x) = \ldots = \lambda_m f_m(x) = 0$, x is optimal.

  Proof.

  - $f_i(x) \neq 0 \Longrightarrow \lambda_i = 0$
  - $f_i(x) = 0 \Longrightarrow \nabla f_i(x)^t(y - x) \leq 0$ for any feasible y
    - if $f(x_i) = 0$, then $\forall y \in X \Longrightarrow f_i(y) \leq f_i(x) \Longrightarrow \nabla f_i^t(x)(y - x) \leq 0$
    - no more proof...
  - for any feasible y, $\nabla f(x)^t(y - x) = -\sum_{i=1}^{m} \lambda_i \nabla f_i(x)^t(y - x) \geq 0$
  - x is optimal.

  □

- the converse is false.

Remark 3.9. COP over nonnegative orthant, with differentiable f

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & x \succeq 0 \end{cases}$$

- $x \in \mathrm{dom}(f)$ is optimal iff

$$x \succeq 0, \quad \begin{cases} \nabla f(x)_i \geq 0 & \text{if } x_i = 0 \\ \nabla f(x)_i = 0 & \text{if } x_i > 0 \end{cases}$$

- *Proof. by observe.*
  - x is optimal $\Longrightarrow \nabla f(x)^t(y - x) \geq 0$ holds **for all feasible** y
  - for $x_i > 0 \Longrightarrow y_i - x_i$ can be positive or negative, so $\nabla f(x)_i = 0$

- for $x_i = 0 \Longrightarrow y_i - x_i \geq 0, \nabla f(x)_i \geq 0$

$\square$

Remark 3.10.

$$\lim_{\varepsilon \to 0} \frac{f(x+a\varepsilon) - f(x)}{\varepsilon} = \langle \nabla f(x), a \rangle, \qquad \varepsilon \to x + a\varepsilon \to f(x+a\varepsilon)$$

Remark 3.11. COP over a simplex, with differentiable f

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & x \succeq 0, \sum_{i=1}^n x_i = 1 \end{cases}$$

- $x \in \mathrm{dom}(f)$ is optimal iff

$$\partial_j f(x) \geq \partial_i f(x) \text{ for all } 1 \leq j \leq n \text{ when } x_i > 0$$

- *Proof.* by observe
  - if x is optimal, $\nabla f(x)^t (y - x) \geq 0$ holds for all feasible y
  - for $x_i > 0 \Longrightarrow y_i - x_i$ can be positive or negative, so $\partial_i f(x) = 0$
  - for $x_i = 0 \Longrightarrow (y_i - x_i) \geq 0, \partial_i f(x) \geq 0$
  - $\partial_i f(x)$ is a constant C for all $x_i > 0$, and $\partial_j f(x) \geq C$ for all $x_j = 0$.
  - the converse is obvious

$$\nabla f(x)^t (y - x) = \sum_{x_i > 0} \partial_i f(x)(y_i - x_i) + \sum_{x_i = 0} \partial_i f(x)(y_i - x_i) \geq C \sum_{i=1}^n (y_i - x_i) = 0$$

$\square$

4 拉格朗日乘子

Remark 4.1. The Lagrange multiplier only tells the properties satisfied by the solution. The solution can not always be obtained by Lagrange multiplier.

Remark 4.2. Standard form optimization problem (not necessarily convex)

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & f_i(x) \leq 0, \quad 1 \leq i \leq m \\ & h_i(x) = 0, \quad 1 \leq i \leq p \end{cases}$$

$x \in \mathcal{D} = \mathrm{dom}(f) \cap (\cap_{i=1}^m \mathrm{dom}\,(f_i)) \cap (\cap_{i=1}^p \mathrm{dom}\,(h_i))$, optimal value denoted p^*. $x \in \mathcal{D}$ **do not need to satisfy constraints.**

- Lagrange function, $L : \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p \mapsto \mathbb{R}$

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x)$$

- Lagrange dual function, $g : \mathbb{R}^m \times \mathbb{R}^p \mapsto \mathbb{R}$

$$g(\lambda, \mu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \mu)$$

  - g is concave, can be $-\infty$ for some (λ, μ)
  - lower bound property: if $\lambda \succeq 0$, then $g(\lambda, \mu) \leq p^*$
    * $L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x) \leq f(x)$
    * $g(\lambda, \mu) = \inf\limits_{x \in \mathcal{D}} L(x, \lambda, \mu) \leq L(x^*, \lambda, \mu) \leq f(x^*) = p^*$.
    * $\max g(\lambda, \mu) = d^*$ with $\lambda \succeq 0$, get $d^* \leq p^*$.

Remark 4.3. Equality constrained norm minimization, with any norm $\|\cdot\|$ of $\mathbb{R}^n$

$$\begin{cases} \text{minimize} & \|x\| \\ \text{subject to} & Ax = b \end{cases}$$

- Lagrangian: $L(x, \mu) = \|x\| + \mu^t (Ax - b)$
- dual function:

$$\begin{aligned} g(\mu) = \inf_{x \in \mathcal{D}} \left(\|x\| + \mu^t Ax - b^t \mu \right) &= \inf_{x \in \mathcal{D}} \left(\|x\| (1 + \mu^t A \frac{x}{\|x\|}) - b^t \mu \right) \\ &= \begin{cases} -b^t \mu, & \text{if } \|A^t \mu\|_* \leq 1 \\ -\infty, & \text{otherwise} \end{cases} \end{aligned}$$

  - $\frac{x}{\|x\|}$ is vector of norm 1.
  - $\|v\|_* = \sup_{\|u\| \leq 1} u^t v$ is the dual norm of $\|\cdot\|$

- lower bound property: $p^* \geq -b^t \mu$ if $\|A^t u\|_* \leq 1$.

Remark 4.4. Standard form LP:

$$\begin{cases} \text{minimize} & c^t x \\ \text{subject to} & x \succeq 0 \\ & Ax = b \end{cases}$$

- Lagrangian:

$$L(x, \lambda, \mu) = c^t x - \lambda^t x + \mu^t (Ax - b) = -\mu^t b + (c + A^t \mu - \lambda)^t x$$

- dual function:

$$
\begin{aligned}
g(\lambda, \mu) &= \inf_{x \in \mathcal{D}} \left(-\mu^t b + (c + A^t\mu - \lambda)^t x\right) \\
&= \begin{cases} -\mu^t b, & \text{if } c + A^t\mu - \lambda = 0 \\ -\infty, & \text{otherwise} \end{cases}
\end{aligned}
$$

- lower bound property: $p^* \geq -\mu^t b$ if $c + A^t\mu \succeq 0$.

Remark 4.5. Tow-way partitioning, for $W \in \mathbb{S}^n$

$$
\begin{cases} \text{minimize} & x^t W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \dots, n \end{cases}
$$

- a nonconvex problem, feasible set contains 2^n discrete points
- Lagrangian: $L(x, \mu) = f(x) + \sum_{i=1}^n \mu_i (x_i^2 - 1)$
- dual function:

$$
\begin{aligned}
g(\mu) &= \inf_{x \in \mathcal{D}} \, (x^t W x + \sum_{i=1}^n \mu_i (x_i^2 - 1)) \\
&= \inf_{x \in \mathcal{D}} \, x^t (W + \operatorname{diag}(\mu)) x - \sum_{i=1}^n \mu_i \\
&= \begin{cases} -\sum_{i=1}^n \mu_i, & \text{if } W + \operatorname{diag}(\mu) \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}
\end{aligned}
$$

- lower bound property: $p^* \geq -\sum_{i=1}^n \mu_i$ if $W + \operatorname{diag}(\mu) \succeq 0$.
  - $p^* \geq n\lambda_{\min}(W)$, where $\lambda_{min}(W)$ is the smallest eigenvalue of W.
  - *Proof.* $W + \operatorname{diag}(\mu) = P^t Q P + \theta I = P^t (Q + \theta I) P \succeq 0$, let $\operatorname{diag}(\mu) = \theta I$, so $\lambda_i + \theta \geq 0, \theta = -\lambda_{\min}(W)$, $p^* \geq n\lambda_{\min}(W)$. □

Remark 4.6. Lagrange dual problem

$$
\begin{cases} \text{maximize} & g(\lambda, \mu) \\ \text{subject to} & \lambda \succeq 0 \end{cases}
$$

- COP, optimal value denoted d^*
- finds best lower bound on $p*$, obtained from Lagrange dual function.
- week duality: $d^* \leq p^*$

- always holds (for both convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

- strong duality: $d^* = p^*$
  - does not hold in general, but usually holds for COP
  - an example that the strong duality does not hold

$$\begin{cases} \text{minimize} & e^{-x} \\ \text{subject to} & x^2/y \leq 0 \end{cases}$$

    * $\mathcal{D} = \{(x,y) : x \in \mathbb{R}, y > 0\}, g(\lambda) = 0 \Longrightarrow d^* = 0 < p^* = 1$

- the dual of dual is primal.
- if the strong duality holds(i.e., $d^* = p^*$), for the primal optimal x^* and the dual optimal (λ^*, μ^*), we have

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

  - x^* minimizes $L(x, \lambda^*, \mu^*)$ on $\mathcal{D}$
  - complementary slackness: $\lambda_i^* f_i(x^*) = 0$ for $i = 1, \ldots, m$
    * $\lambda_i^* > 0 \Longrightarrow f_i(x^*) = 0$
    * $f_i(x^*) < 0 \Longrightarrow \lambda_i^* = 0$

Remark 4.7. KKT conditions:

- primal constraints: $f_1(x) \leq 0, \ldots, f_m(x) \leq 0, h_1(x) = 0, \ldots, h_p(x) = 0$
- dual constraints: $\lambda \succeq 0$
- complementary slackness: $\lambda_i f_i(x) = 0$ $(i = 1, \ldots, m)$
- gradient of Lagrangian with respect to x vanishes: $\partial_x L(x, \lambda, \mu) = 0$

KKT gave information about (x^*, λ^*, μ^*). In some cases, the optimal solution can be obtained through these four conditions.

Remark 4.8. **Important conclusion**

- With strong duality:

$$(x^*, \lambda^*, \mu^*) \text{ optimal } \Longrightarrow (x^*, \lambda^*, \mu^*) \text{ satisfy KKT}$$

- For COP, with strong duality:

$$(x^*, \lambda^*, \mu^*) \text{ optimal } \Longleftrightarrow (x^*, \lambda^*, \mu^*) \text{ satisfy KKT}$$

Remark 4.9. Convex optimization problem

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & Ax = b \end{cases}$$

- Slater's constraint qualifiction: strong duality holds if

$$\exists x \in \textbf{int}\mathcal{D} \quad \text{s.t. } Ax = b \text{ and } f_i(x) < 0, \quad i = 1, \dots, m$$

  - x is a interior point of $\mathcal{D}$, called a strictly feasible point.
  - also guarantees that the dual optimum is attained if $p^* > -\infty$, i.e., there exists (λ^*, μ^*) s.t. $g(\lambda^*, \mu^*) = d^* = p^*$. (Prove difficulty)
    - minimize e^{-x} has a minimum, but can not attain the optimal point.

- Slater's weak constraint qualification: if the first k constraint functions $f_1, \dots, d_k$ are affine, the strong duality holds if

$$\exists x \in \textbf{int}\mathcal{D} \quad \text{s.t. } Ax = b, f_1(x), \dots, f_k(x) \leq 0, f_{k+1}(x), \dots, f_m(x) < 0$$

  - strong duality holds for any LP/QP if it is feasible.

- if Slater's(weak) condition is satisfied, the strong duality holds, and

$$(x^*, \lambda^*, \mu^*) \text{ optimal } \Longleftrightarrow (x^*, \lambda^*, \mu^*) \text{ satisfy KKT}$$

  - generalizes optimality condition $\nabla f(x^*) = 0$ for unconstrained problem

Remark 4.10. **Important conclusion**

COP, f, f_i differentiable:

- strong duality: optimal $\Longrightarrow$ KKT

- COP: KKT $\Longrightarrow$ optimal
- COP + strong duality: optimal $\Longleftrightarrow$ KKT
- COP + Slater(weak Slater) or affine constraints, LP/QP
  - optimal $\Longleftrightarrow$ KKT
  - x^* optimal $\Longleftrightarrow$ there exists (λ^*, μ^*) s.t. (x^*, λ^*, μ^*) satisfy KKT
  - x^* optimal $\Longrightarrow$ x^* minimizes $L(x, \lambda^*, \mu^*)$ on $\mathcal{D}$ for some (λ^*, μ^*), i.e.

$$x^* = \arg\min_{x \in \mathcal{D}} \left(f(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + (\mu^*)^t (Ax - b) \right)$$

Remark 4.11. COP over nonnegative orthant, with differentiable f

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & x \succeq 0 \end{cases}$$

- COP + addine constraints
- Lagrange function: $L(x, \lambda) = f(x) - \lambda^t x$
- x is optimal iff there exists $\lambda \in \mathbb{R}^n$ s.t. (x, λ) satisfy
  - $x \succeq 0$
  - $\lambda \succeq 0$
  - $\lambda_i x_i = 0$
  - $\partial_x L(x, \lambda) = 0 \Longleftrightarrow \nabla f(x) = \lambda$
- optimality conditions: $x \in \mathrm{dom}(f)$ is optimal iff

$$x \succeq 0, \nabla f(x) \succeq 0, \text{ and } x_i \partial_i f(x) = 0 \text{ for each } i \in \{1, \ldots, n\}$$

Remark 4.12. COP over a simplex, with differentiable f

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & x \succeq 0, \sum_{i=1}^n x_i = 1 \end{cases}$$

- COP + affine constraints
- Lagrange function: $L(x, \lambda, \mu) = f(x) - \lambda^t x + \mu(\sum_{i=1}^n x_i - 1)$
- x is optimal iff there exist $\lambda \in \mathbb{R}^n$ s.t. (x, λ, μ) satisfy

  - $x \succeq 0$ and $\sum_{i=1}^{n} = 1$
  - $\lambda \succeq 0$
  - $\lambda_i x_i = 0$
  - $\partial_x L(x, \lambda, \mu) = 0 \iff \nabla f(x) + \mu \mathbf{1} = \lambda$

- $\begin{cases} x_i \neq 0 \implies \lambda_i = 0 \implies \nabla_i f(x) = -\mu \\ x_i = 0 \implies \nabla_i f(x) \geq -\mu \end{cases}$

- optimality conditions: $x \in \mathrm{dom}(f)$ is optimal iff $x \succeq 0, \sum_{i=1}^{n} x_i = 1$, $\partial_i f(x)$ is a constant(noted C) for each $x_i > 0$, and $\partial_i f(x) \geq C$ holds for each $x_i = 0$.

- seems $C = \mu = 0$?

Remark 4.13. Equality constrained COP, with differentiable f, $A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p$

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{cases}$$

- COP + affine constraints

- Lagrange function: $L(x, \mu) = f(x) + \mu^t (Ax - b)$

- x is optimal iff there exists $\mu \in \mathbb{R}^p$ s.t. (x, μ) satisfy
  - $Ax = b$
  - $\partial_x L(x, \mu) = 0 \iff \nabla f(x) + A^t \mu = 0$

- optimality conditions: $x \in \mathrm{dom}(f)$ is optimal iff

$$Ax = b, \text{ and there exists } \mu \in \mathbb{R}^p \text{ s.t. } \nabla f(x) + A^t \mu = 0$$

Remark 4.14. Least-norm solution of linear eauations

$$\begin{cases} \text{minimize} & \|x\|^2 \\ \text{subject to} & Ax = b \end{cases}$$

- COP + affine constraints

- Lagrangian: $L(x, \mu) = \|x\|^2 + \mu^t (Ax - b)$

- x is optimal $\iff$ there exists μ s.t. (x, μ) satisfy: $Ax = b, 2x + A^t \mu = 0 \iff AA^t \mu = -2b, 2x + A^t \mu = 0$

- for any $z \in \mathbb{R}^n$, we take $z = u + v$ with $u \in \mathcal{N}(A)$ and $v \in \mathcal{R}(A^t)$, then $Az = Av$, and thus $\mathcal{R}(A) = \mathcal{R}(AA^t)$
- $\exists \mu$ s.t. $AA^t\mu = -2b \iff b \in \mathcal{R}(AA^t) \iff b \in \mathcal{R}(A) \iff Ax = b$ has at least one solution $\iff$ the problem is feasible
- if the problem is feasible, $AA^t\theta = b \implies x = A^t\theta$ is optimal.

Remark 4.15. Water-filling, with each $\alpha_i > 0$

$$\begin{cases} \text{minimize} & -\sum_{i=1}^n \log(x_i + \alpha_i) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^t x = 1 \end{cases}$$

- COP + affine constraints
- Lagrange function: $L(x, \lambda, \mu) = -\sum_{i=1}^n \log(x_i + \alpha_i) - \lambda^t x + \mu(\mathbf{1}^t x - 1)$
- x is optimal $\iff$ there exists (λ, μ) s.t. (x, λ, μ) satisfy

$$x \succeq 0, \quad \mathbf{1}^t x = 1, \quad \lambda \succeq 0, \quad \lambda_i x_i = 0, \quad \frac{1}{x_i + \alpha_i} + \lambda_i = \mu$$

  - the objective function is continues on a compact set, the minimizer exists
  - if $\mu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\mu - \alpha_i$
  - if $\mu \geq 1/\alpha_i$: $\lambda_i = \mu - 1/\alpha_i$ and $x_i = 0$
  - $x_i = \max\{0, 1/\mu - \alpha_i\}$
  - determine μ from $\mathbf{1}^t x = \sum_{i=1}^n \max\{0, 1/\mu - \alpha_i\} = 1$

Remark 4.16. Entropy maximization

$$\begin{cases} \text{minimize} & \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & \mathbf{1}^t x = 1, Ax \preceq b \end{cases}$$

- COP + affine constraints
- Lagrangian: $L(x, \lambda, \mu) = \sum_{i=1}^n x_i \log x_i + \lambda^t(Ax - b) + \mu(\mathbf{1}^t x - 1)$
- x is optimal $\iff$ there exists (λ, μ) s.t. (x, λ, μ) satisfy

$$\mathbf{1}^t x = 1, \quad Ax \preceq b, \quad \lambda \succeq 0, \quad \lambda_i (Ax - b)_i = 0, \quad 1 + \log x_i + a_i^t \lambda + \mu = 0$$

  - $A = (a_1, \ldots, a_n)$, x is optimal $\implies x_i = e^{-1 - a_i^t\lambda - \mu} = \frac{e^{-a_i^t\lambda}}{\sum_{k=1}^n e^{-a_k^t\lambda}}$

  - x is optimal $\Longrightarrow \begin{cases} \text{minimize} & \sum_{i=1}^n \frac{e^{-a_i^t\lambda}}{\sum_{k=1}^n e^{-a_k^t\lambda}} \left(-a_k^t\lambda - \log\left(\sum_{k=1}^n e^{-a_k^t\lambda}\right)\right) \\ \text{subject to} & \lambda \succeq 0, \quad \sum_{i=1}^n a_i e^{-a_i^t\lambda} \preceq \left(\sum_{k=1}^n e^{-a_k^t\lambda}\right) b \end{cases}$
  - KKT failure.

- dual function: $g(\lambda,\mu) = -b^t\lambda - \mu - e^{-\mu-1}\sum_{i=1}^n e^{-a_i^t\lambda}$
- dual problem: $\begin{cases} \text{minimize} & b^t\lambda + \log\left(\sum_{i=1}^n e^{-a_i^t\lambda}\right) \\ \text{subject to} & \lambda \succeq 0 \end{cases}$
  - a better way!

Remark 4.17. Minimizing a separable function subject to an equality constraint, with $f_i : \mathbb{R} \mapsto \mathbb{R}$ differentiable and strictly convex

$$\begin{cases} \text{minimize} & f(x) = \sum f_i(x_i) \\ \text{subject to} & a^t x = b \end{cases}$$

- COP + affine constraints
- Lagrange function: $L(x,\mu) = \sum f_i(x_i) + \mu(a^t x - b)$
- x is optimal iff $\exists \mu \in \mathbb{R}$ s.t. (x,μ) satisfy: $a^t x = b, f_i'(x_i) + \mu a_i = 0$
  - seems nothing!
- dual function: $g(\mu) = -b\mu - \sum f_i^*(-\mu a_i)$
- dual problem: minimize$(b\mu + \sum f_i^*(-\mu a_i))$
  - unconstrained problem with $\mu \in \mathbb{R}$, much simpler!

5 凸优化应用

5.1 SVM

Remark 5.1. For given $(x_1,y_1),\ldots,(x_m,y_m), \quad x_i \in \mathbb{R}^n, y_i \in \{1,-1\}$, suppose that they are separable, i.e., there exists $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ s.t.

$$\begin{cases} y_i = 1 & \text{if } a^t x_i + b > 0 \\ y_i = -1 & \text{if } a^t x_i + b < 0 \end{cases}$$

- In this separable case, we can suppose further that

$$\begin{cases} y_i = 1 & \text{if } a^t x_i + b \geq 1 \\ y_i = -1 & \text{if } a^t x_i + b \leq -1 \end{cases}$$

  $y_i(a^t x_i + b) \geq 1$, for each $i \in \{1, \ldots, m\}$

- The distance between the two hyperplanes $\{x : a^t x + b = 1\}$ and $\{x : a^t x + b = -1\}$ is $\frac{2}{\|a\|}$.

- To separate the data set by the two hyperplanes with maximum margin, we then consider the following QP

$$\begin{cases} \text{minimize} & \frac{1}{2}\|a\|^2 \\ \text{subject to} & y_i(a^t x_i + b) \geq 1, \quad 1 \leq i \leq m \end{cases}$$

Remark 5.2. For the non-separable case, we may consider the following optimization problem, where $C > 0$ is a given constant

$$\text{minimize} \frac{1}{2}\|a\|^2 + C \sum_{i=1}^{m} l_{0/1}(y_i(a^t x_i + b) - 1)$$

- 0/1-loss: $l_{0/1}(z) = \begin{cases} 1, & \text{if } z < 0 \\ 0, & \text{otherwise} \end{cases}$

- however, $l_{0/1}$ is non-convex.

Then we consider the following optimization problem, where $C > 0$ is a given constant

$$\text{minimize} \quad \frac{1}{2}\|a\|^2 + C \sum_{i=1}^{m} \max\left(1 - y_i\left(a^t x_i + b\right), 0\right)$$

which is equivalent to the following QP

$$\begin{cases} \underset{(a,b,\xi)}{\text{minimize}} & \frac{1}{2}\|a\|^2 + C \sum_{i=1}^{m} \xi_i \\ \text{subject to} & y_i\left(a^t x_i + b\right) \geq 1 - \xi_i, \quad 1 \leq i \leq m \\ & \xi \succeq 0 \end{cases}$$

- Lagrange function:

$$L(a, b, \xi, \alpha, \beta) = \frac{1}{2}\|a\|^2 + C \sum \xi_i + \sum \alpha_i \left(1 - \xi_i - y_i\left(a^t x_i + b\right)\right) - \sum \beta_i \xi_i$$

- dual function: $g(\alpha, \beta) = \inf_{(a,b,\xi)} L(a, b, \xi, \alpha, \beta)$

$$g(\alpha, \beta) = -\frac{1}{2}\left\|\sum \alpha_i y_i x_i\right\|^2 + \sum \alpha_i \quad \text{if } \sum \alpha_i y_i = 0 \text{ and } C = \alpha_i + \beta_i$$

- Dual problem(QP)

$$\begin{cases} \underset{(\alpha,\beta)}{\text{maximize}} & -\frac{1}{2}\left\|\sum \alpha_i y_i x_i\right\|^2 + \sum \alpha_i \\ \text{subject to} & \alpha \succeq 0, \beta \succeq 0 \\ & \sum \alpha_i y_i = 0,\ C = \alpha_i + \beta_i \end{cases}$$

- it can be further simplified

- Dual problem(QP)

$$\begin{cases} \underset{\alpha}{\text{minimize}} & \frac{1}{2}\left\|\sum \alpha_i y_i x_i\right\|^2 - \sum \alpha_i \\ \text{subject to} & 0 \le \alpha_i \le C, \quad 1 \le i \le m \\ & \sum \alpha_i y_i = 0 \end{cases}$$

  - much simpler, e.g., by SMO(cequential minimal optimization)
  - by KKT, $\partial_a L = 0 \Longrightarrow a = \sum \alpha_i y_i x_i$
  - by KKT, complementary slackness: if $0 < \alpha_i < C$, then $\beta_i > 0$, and thus

$$\begin{cases} \alpha_i(1 - \xi_i - u_i(a^t x_i + b)) = 0 \\ \beta_i \xi_i = 0 \end{cases} \Longrightarrow 1 - y_i\left(a^t x_i + b\right) = 0 \Longrightarrow b = y_i - a^t x_i$$

5.2 C-means (K-means) clustering

Remark 5.3. $x_1, \ldots, x_N \in \mathbb{R}^n$ are given

$$\begin{cases} \text{minimize} & \sum_{i=1}^N \sum_{k=1}^M u_{i,k} \|x_i - y_k\|^2 \\ \text{subject to} & \sum_{k=1}^M u_{i,k} = 1, \quad u_{i,k} \in \{0,1\} \end{cases}$$

- non-convex on (x, y)

- two-steps optimization algorithm

- when $u_{i,k}$ are given, it is an unconstrained QP on y_k
  - $\partial_{y_k} = 0 \Longrightarrow \sum_{i=1}^N u_{i,k}\left(y_k - x_i\right) = 0 \Longrightarrow y_k = \frac{\sum_{i=1}^N u_{i,k} x_i}{\sum_{i=1}^N u_{i,k}}$

- when y_k are given, it is simple that $u_{i,k} = 1$ iff

$$k = \underset{1 \le j \le M}{\arg\min} \left\|x_i - y_j\right\|^2$$

5.3 Fuzzy C-means clustering

Remark 5.4. $x_1, \ldots, x_N \in \mathbb{R}^n$ are given, and $\alpha > 1$

$$\begin{cases} \text{minimize} & \sum_{i=1}^N \sum_{k=1}^M u_{i,k}^\alpha \left\|x_i - y_k\right\|^2 \\ \text{subject to} & \sum_{k=1}^M u_{i,k} = 1, \quad \forall i \in \{1, \ldots, N\} \end{cases}$$

- two-steps optimization algorithm
- when $u_{i,k}$ are given, it is an unconstrained QP on y_k
  - $\partial_{y_k} = 0 \Longrightarrow \sum_{i=1}^N u_{i,k}^\alpha (y_k - x_i) = 0 \Longrightarrow y_k = \frac{\sum_{i=1}^N u_{i,k}^\alpha x_i}{\sum_{i=1}^N u_{i,k}^\alpha}$
- when y_k are given, COP + affine constraints
  - Lagrange function:

$$L(u, \mu) = \sum_{i=1}^N \sum_{k=1}^M u_{i,k}^\alpha \left\|x_i - y_k\right\|^2 + \sum_{i=1}^N \mu_i \left(\sum_{k=1}^M u_{i,k} - 1 \right)$$

  - u is optimal iff there exists $\mu \in \mathbb{R}^N$ s.t. (u, μ) satisfy

$$\sum_{k=1}^M u_{i,k} = 1, \quad \alpha u_{i,k}^{\alpha-1} \left\|x_i - y_k\right\|^2 + \mu_i = 0$$

  - for optimal u, $u_{i,k} = \frac{\|x_i - y_k\|^{-\frac{2}{\alpha-1}}}{\sum_{j=1}^M \|x_i - y_j\|^{-\frac{2}{\alpha-1}}}$

5.4 LLE

Remark 5.5. For given data $x_1, \ldots, x_N \in \mathbb{R}^n$, two-step optimization

- assign neighbors to each data point x_i (e.g., by using the K nearest neighbors)
- compute the weights $w_{i,j}$ that best linearly reconstruct x_i from its neighbors
  - minimize $\sum_{i=1}^N \|x_i - \sum_{j=1}^N w_{i,j} x_j\|$ subject to $\sum_{j=1}^N w_{i,j} = 1$
- compute the low-dimensional vectors $y_i \in \mathbb{R}^m$ best reconstructed by $w_{i,j}$
  - minimize $\sum_{i=1}^N \|y_i - \sum_{j=1}^N w_{i,j} y_j\|$ subject to $\mathrm{Cov}(y, y) = I$

Remark 5.6. $x_1, \ldots, x_N, x \in \mathbb{R}^n$ are given

$$\begin{cases} \text{minimize} & \|x - \sum_{i=1}^N w_i x_i\|^2 \\ \text{subject to} & \sum_{i=1}^N w_i = 1 \end{cases}$$

- as $\sum_{i=1}^{N}=1, x-\sum_{i=1}^{N} w_i x_i=\sum_{i=1}^{N}(x-x_i)$, then we have

$$\left\|x-\sum_{i=1}^{N} w_i x_i\right\|^2=w^t Q w, \quad Q_{i,j}=(x-x_i)^t (x-x_j)$$

- Lagrange function: $L(w,\mu)=\frac{1}{2}w^tQw+\mu(\sum_{i=1}^{N} w_i-1)$
- w is optimal iff there exists $\mu\in\mathbb{R}$ s.t. (w,μ) satisfy
  - $\sum_{i=1}^{N} w_i=1$
  - $\partial_w L(w,\mu)=0 \Longleftrightarrow Qw+\mu\mathbf{1}=0 \Longleftrightarrow w=-\mu Q^{-1}\mathbf{1}$
- w is optimal iff $w=\frac{Q^{-1}\mathbf{1}}{\mathbf{1}^tQ^{-1}\mathbf{1}}$

Remark 5.7. $w_{i,j}\in\mathbb{R}$ are given

$$\begin{cases}\text{minimize} & \sum_{i=1}^{N}\|y_i-\sum_{j=1}^{N} w_{i,j}y_j\|^2\\ \text{subject to} & \sum_{i=1}^{N} y_i y_i^t=NI\end{cases}$$

- $\sum_{i=1}^{N}\left\|y_i-\sum_{j=1}^{N} w_{i,j}y_j\right\|^2=\|Y-WY\|^2=\operatorname{tr}(Y^tMY)$
  - $Y=(y_1,\ldots,y_N)^t, W=(w_{i,j}), M=(I-W)^t(I-W)$
- Lagrange function: $L(Y,\mu)=\operatorname{tr}(T^tMY)+\operatorname{tr}\big(\mu(Y^Y-NI)\big)$
- Y is optimal $\Longrightarrow$ there exists μ s.t. Y,μ satisfy

$$\sum_{i=1}^{N} y_i y_i^t=NI, \quad 2MY+Y\left(\mu+\mu^t\right)=0$$

  - $-\frac{\mu+\mu^t}{2}=PQP^t, 2MY+Y(\mu+\mu^t)=0 \Longleftrightarrow M(YP)=(YP)Q$
  - $YP\triangleq\tilde{Y} \Longrightarrow \operatorname{tr}(Y^tMY)=\operatorname{tr}\left(\tilde{Y}^tM\tilde{Y}\right)=\operatorname{tr}\left(\tilde{Y}^t\tilde{Y}Q\right)=N\operatorname{tr}(Q)$
- if Y is composed of the normalized eigenvectors corresponding to the smallest m eigenvalues of M, it is optimal
  - practically, the eigenvector $\mathbf{1}=(1,\ldots,1)^t$ is discarded.

6 无约束凸优化问题求解

7 有约束凸优化问题求解