Zerrad
Avigaïl
(MASH)

Convex optimization - exam

n°=1

The program $\min \|x\|_\infty$ s.t $Ax = b$ is equivalent to $\min t$ s.t $-t\mathbb{1} \leq x \leq t\mathbb{1}$, $Ax = b$

We can rewrite the constraint $-t\mathbb{1} \leq x \leq t\mathbb{1}$ as $\begin{pmatrix} x \\ -x \end{pmatrix} \leq t\mathbb{1}_{2n}$.

Then, if we set

$$X = \begin{pmatrix} x \\ -x \end{pmatrix} \; ; \quad \tilde{A} = \begin{pmatrix} A & (0) \\ (0) & A \end{pmatrix} \quad \text{and} \quad \tilde{b} = \begin{pmatrix} b \\ -b \end{pmatrix},$$

we can rewrite the program as following: $\min t$ s.t $X \leq t\mathbb{1}_{2n}$, $\tilde{A}X = \tilde{b}$, which is a LP.

Lets compute its dual.

Lagrangian: $L(X,t,\lambda,\nu) = t + \lambda^T(X - t\mathbb{1}_{2n}) + \nu^T(\tilde{b} - \tilde{A}X)$

Lagrange dual function: $g(\lambda,\nu) = \inf_{X,t} L(X,t,\lambda,\nu)$

$$= \inf_{X,t} (1 - \lambda^T\mathbb{1}_{2n})t + (\lambda - \tilde{A}^T\nu)^T X + \nu^T\tilde{b}$$

$$= \inf_t (1 - \lambda^T\mathbb{1}_{2n})t + \inf_X (\lambda - \tilde{A}^T\nu)^T X + \nu^T\tilde{b}$$

$$= \begin{cases} \nu^T\tilde{b} & \text{if } \lambda - \tilde{A}^T\nu = 0 \text{ and } 1 = \lambda^T\mathbb{1}_{2n} \\ -\infty & \text{otherwise} \end{cases}$$

Dual problem: $\max g(\lambda,\nu)$ s.t $\lambda \geq 0$ ie $\max \nu^T\tilde{b}$ s.t $\tilde{A}^T\nu \geq 0$, $\nu^T\tilde{A}\mathbb{1}_{2n} = 1$

$n° = 2$

First, we will show that the function f is convex.

Since f is twice differentiable, lets compute its hessian.

$$\nabla f = \begin{pmatrix} 1 + \log v_1 \\ \vdots \\ 1 + \log v_n \end{pmatrix} \quad ; \quad H_f = \begin{pmatrix} 1/v_1 & & (0) \\ & \ddots & \\ (0) & & 1/v_n \end{pmatrix} \succ 0 \text{ on } \mathbb{R}^n_{++}$$

Since $H_f \succ 0$, we have that f is strictly convex on $\mathbb{R}^n_{++}$.

Then, by definition, we have:

$$\forall u, v \in \mathbb{R}^n_{++} \text{ such that } u \neq v, \; f(u) > f(v) + \nabla f(v)^T (u - v)$$
$$\Rightarrow f(u) - f(v) - \nabla f(v)^T (u - v) > 0$$

We deduce that $D_{kl}(u, v) \geq 0 \;\; \forall u, v \in \mathbb{R}^n_{++}$.

Moreover, $D_{kl}(u, v) = 0 \iff u = v$. Indeed, $\forall u \neq v$, we have $f(u) > f(v) + \nabla f(v)^T (u - v)$ so $f(u) = f(v) + \nabla f(v)^T (u - v)$ if and only if $u = v$.

$n° = 3$

* First, we rewrite this program as a SOCP.

We focus on rewriting the constraints $x^T (A - bb^T) x \leq 0$

$b^T x \geq 0$

Since $A \succeq 0$, we have $A = V^T V$ for some $V \in S_n$. We then have:

$$\begin{aligned} x^T (A - bb^T) x &= x^T (V^T V - bb^T) x \\ &= (x^T V^T V - x^T bb^T) x \\ &= x^T V^T V x - x^T bb^T x \\ &= (Vx)^T (Vx) - (b^T x)^T (b^T x) \\ &= \|Vx\|_2^2 - \|b^T x\|_2^2 . \end{aligned}$$

Since $b^T x \in \mathbb{R}$, we have $\|b^T x\|_2^2 = (b^T x)^2$.

The constraints $\begin{array}{l} x^T(A - bb^T x) \le 0 \\ b^T x \ge 0 \end{array}$ become $\begin{array}{l} \|Vx\|_2^2 \le (b^T x)^2, \\ b^T x \ge 0 \end{array}$

and that is equivalent to $\|Vx\|_2 \le b^T x$.

The given program can be written as:
$$\begin{array}{ll} & \min c^T x \\ \text{s.t} & \|Vx\|_2 \le b^T x \\ & Dx = g \end{array}$$

This is a SOCP, which is by definition a convex optimization problem.

* We introduce new variables $y = Vx$ and $t = b^T x$.

We then want to compute the dual of the following:
$$\begin{array}{ll} & \min c^T x \\ \text{s.t} & \|y\|_2 \le t \\ & y = Vx \\ & t = b^T x \\ & Dx = g \end{array}$$

<u>Lagrangian</u>: $L(x, y, t, \lambda, \nu, \mu, \pi) = c^T x + \lambda^T(\|y\|_2 - t) + \nu^T(y - Vx) + \mu^T(t - b^T x) + \pi^T(g - Dx)$

$$= (c - V^T\nu - b\mu - D^T\pi)^T x + (\lambda^T \|y\|_2 + \nu^T y) + (-\lambda^T + \mu^T)t + \pi^T g$$

<u>Lagrange dual function</u>: $g(\lambda, \nu, \mu, \pi) = \inf_{x,y,t} L(x, y, t, \lambda, \nu, \mu, \pi)$

$$\Rightarrow g(\lambda, \nu, \mu, \pi) = \inf_x (c - V^T\nu - b\mu - D^T\pi)^T x + \inf_y (\lambda^T \|y\|_2 + \nu^T y) + \inf_t (-\lambda^T + \mu^T)t + \pi^T g$$

$\hookrightarrow \inf_x (c - V^T\nu + b\mu - D^T\pi)^T x = \begin{cases} 0 & \text{if } c - V^T\nu - b\mu - D^T\pi = 0 \\ -\infty & \text{otherwise} \end{cases}$

$\hookrightarrow \inf_y \lambda^T \|y\|_2 + \nu^T y = \begin{cases} 0 & \text{if } \|\nu\|_2 \le \lambda \\ -\infty & \text{otherwise} \end{cases}$

$\hookrightarrow \inf_t (-\lambda^T + \mu^T)t = \begin{cases} 0 & \text{if } \lambda = \mu \\ -\infty & \text{otherwise} \end{cases}$

$$\Rightarrow g(\lambda, \nu, \mu, \pi) = \begin{cases} \pi^T g & \text{if } c - V^T\nu - b\mu - D^T\pi = 0,\ \|\nu\|_2 \le \lambda \text{ and } \lambda = \mu \\ -\infty & \text{otherwise} \end{cases}$$

Dual problem: $\max \ \pi^T g$

$$\text{s.t} \quad \lambda \geqslant \|\nu\|_2$$

$$c - V^T\nu - b\lambda - \Delta^T\pi = 0$$

n°=4

We introduce new variables $y_i = b_i - a_i^T x$.

We rewrite the problem as :

$$\min - \sum_{i=1}^{m} \log(y_i)$$

$$\text{s.t} \quad y = b - Ax$$

with A such that a_i^T is its i-th row.

Lagrangian: $L(x, y, \mu) = - \sum_{i=1}^{m} \log(y_i) + \mu^T(y - b + Ax)$

Lagrange dual function: $g(\mu) = \inf_{x,y} L(x, y, \mu)$

$$= \inf_{x,y} - \sum_{i=1}^{m} \log(y_i) + \mu^T y + \mu^T Ax - \mu^T b$$

$$= \inf_x (\mu^T Ax) + \inf_y \left(- \sum_{i=1}^{m} \log(y_i) + \mu^T y\right) - \mu^T b$$

↳ $\inf_x (\mu^T Ax) = \begin{cases} 0 & \text{if } \mu^T A = 0 \\ -\infty & \text{otherwise} \end{cases}$

↳ $\inf_y \left(- \sum_{i=1}^{m} \log(y_i) + \mu^T y\right)$. Since $y \mapsto - \sum_{i=1}^{m} \log(y_i) + \mu^T y$ is convex, we have,

by setting the gradient at 0 : $\forall i = 1, \dots, m, \ -\frac{1}{y_i} + \mu_i = 0 \iff y_i = \frac{1}{\mu_i}$ for $\mu_i > 0$

and then $\inf_y \left(- \sum_{i=1}^{m} \log(y_i) + \mu^T y\right) = - \sum_{i=1}^{m} \log\left(\frac{1}{\mu_i}\right) + \sum_{i=1}^{m} \mu_i \times \frac{1}{\mu_i}$ if $\mu > 0$

$\Rightarrow g(\mu) = \sum_{i=1}^{m} \log(\mu_i) + m - b^T\mu$ if $A^T\mu = 0$ and $\mu > 0$; $-\infty$ otherwise

Dual problem : $\max \ \sum_{i=1}^{m} \log(\mu_i) + m - b^T\mu$.

$$\text{s.t} \quad \mu > 0$$

$$A^T\mu = 0$$

n° 5

The auxiliary function ϕ is differentiable because f_0 and $x \mapsto \alpha \|Ax - b\|_2^2$ are differentiable. Since $\tilde{x}$ is a minimizer of ϕ, we have that $\nabla\phi(\tilde{x}) = 0$

$$\iff \nabla f_0(\tilde{x}) + 2\alpha A^T (A\tilde{x} - b) = 0.$$

We can deduce that $\tilde{x}$ is also a minimizer of $h : x \mapsto f_0(x) + \nu^T(Ax - b)$

where $\nu := 2\alpha(A\tilde{x} - b)$.

Indeed, $\nabla h(x) = \nabla f_0(x) + A^T \nu = \nabla f_0(x) + A^T \times 2\alpha(A\tilde{x} - b)$.

The Lagrangian function associated to the problem (1) is $L(x, \nu) = f_0(x) + \nu^T(Ax - b)$

The Lagrange dual function is then : $g(\nu) = \inf_x L(x, \nu)$

$$= \inf_x f_0(x) + \nu^T(Ax - b)$$

$$= f_0(\tilde{x}) + \nu^T(A\tilde{x} - b)$$

$$= f_0(\tilde{x}) + 2\alpha \|A\tilde{x} - b\|_2^2 .$$

We deduce that $\nu = 2\alpha(A\tilde{x} - b)$ is a dual feasible point for (1).

By the lower bound property, we also deduce that $f_0(x^*) \geq g(\nu)$ where $f_0(x^*)$ is the optimal value of (1).

Then, we have $f_0(x^*) \geq f_0(\tilde{x}) + 2\alpha \|A\tilde{x} - b\|_2^2$.

n°=6

1) Derivation of the dual

Lagrangian: $L(\omega, z, \lambda, \mu) = \frac{1}{2}\|\omega\|_2^2 + C\mathbb{1}^T z + \sum_{i=1}^{m} \lambda_i (1 - z_i - y_i(\omega^T x_i)) - \mu^T z$

$$= \left(\frac{1}{2}\|\omega\|_2^2 - \sum_{i=1}^{m} \lambda_i y_i(\omega^T x_i)\right) + (C\mathbb{1} - \lambda - \mu)^T z + \sum_{i=1}^{m} \lambda_i$$

Lagrange dual function: $g(\lambda,\mu) = \inf_{\omega, z} L(\omega, z, \lambda, \mu)$

$$= \inf_{\omega} \left(\frac{1}{2}\|\omega\|_2^2 - \sum_{i=1}^{m} \lambda_i y_i(\omega^T x_i)\right) + \inf_{z} \left((C\mathbb{1} - \lambda - \mu)^T z\right) + \sum_{i=1}^{m} \lambda_i$$

↳ $\inf_{z} (C\mathbb{1} - \lambda - \mu)^T z = \begin{cases} 0 & \text{if } C\mathbb{1} - \lambda - \mu = 0 \\ -\infty & \text{otherwise} \end{cases}$

↳ $\inf_{\omega} \left(\frac{1}{2}\|\omega\|_2^2 - \sum_{i=1}^{m} \lambda_i y_i(\omega^T x_i)\right)$. By setting the gradient at 0, we have

$$\frac{1}{2} \times 2\omega - \sum_{i=1}^{m} \lambda_i y_i x_i = 0 \iff \omega = \sum_{i=1}^{m} \lambda_i y_i x_i$$

$$\Rightarrow \inf_{\omega} \left(\frac{1}{2}\|\omega\|_2^2 - \sum_{i=1}^{m} \lambda_i y_i(\omega^T x_i)\right) = \frac{1}{2}\left\|\sum_{i=1}^{m} \lambda_i y_i x_i\right\|_2^2 - \left(\sum_{i=1}^{m} \lambda_i y_i x_i\right)^T \sum_{j=1}^{m} \lambda_j y_j x_j$$

$$= -\frac{1}{2}\left\|\sum_{i=1}^{m} \lambda_i y_i x_i\right\|_2^2$$

Then $g(\lambda,\mu) = \begin{cases} \sum_{i=1}^{m} \lambda_i - \frac{1}{2}\left\|\sum_{i=1}^{m} \lambda_i y_i x_i\right\|_2^2 & \text{if } C\mathbb{1} - \lambda - \mu = 0 \\ -\infty & \text{otherwise} \end{cases}$

Dual problem: $\max g(\lambda,\mu)$ s.t. $\lambda \geq 0$, $\mu \geq 0$ ie $\max \mathbb{1}^T\lambda - \frac{1}{2}\left\|\sum_{i=1}^{m} \lambda_i y_i x_i\right\|_2^2$ s.t. $0 \leq \lambda \leq C\mathbb{1}_m$

2) first step: write problems as QP.

Primal problem: We want to write the primal problem into a QP, of the form

$$\min \ v^T Q v + p^T v$$
$$\text{s.t} \quad Av \leq b$$

we have: $\min \ \frac{1}{2} \|w\|_2^2 + C\mathbb{1}^T z$

$$\text{s.t} \quad y_i (w^T x_i) \geq 1 - z_i \qquad i = 1, \dots, m$$
$$z \geq 0$$

We set: $v = \begin{pmatrix} w \\ z \end{pmatrix} \in \mathbb{R}^{n+m}$; $\underset{(n+m)\times(n+m)}{Q} = \frac{1}{2} \begin{pmatrix} I_n & (0)_{n\times m} \\ (0)_{m\times n} & (0)_{m\times m} \end{pmatrix}$; $p = C \begin{pmatrix} (0)_n \\ \mathbb{1}_m \end{pmatrix} \in \mathbb{R}^{n+m}$

We can rewrite the constraints $y_i (w^T x_i) \geq 1 - z_i \ \ i = 1, \dots, m$ as

$$\mathbb{1} - z - YX^T w \leq 0 \quad , \text{i.e} \ -z - YX^T w \leq -\mathbb{1}$$

Then, we set: $\underset{(2m)\times(n+m)}{A} = \begin{pmatrix} -YX^T & -I_m \\ (0)_{m\times n} & -I_m \end{pmatrix}$ and $b = \begin{pmatrix} -\mathbb{1}_m \\ (0)_m \end{pmatrix} \in \mathbb{R}^{2m}$

For the initial point $v_0 = \begin{pmatrix} w_0 \\ z_0 \end{pmatrix}$, we will take v_0 such that $Av_0 \leq b$.

Dual problem: Same reasoning

we have: $\min \ \frac{1}{2} \| \sum_{i=1}^{m} \lambda_i y_i x_i \|_2^2 - \mathbb{1}^T \lambda$

$$\text{s.t} \quad 0 \leq \lambda \leq C\mathbb{1}_m$$

We set: $v = \lambda \in \mathbb{R}^m$; $Q = \frac{1}{2} YXX^TY$; $p = -\mathbb{1}_m$

$$A = \begin{pmatrix} I_m \\ -I_m \end{pmatrix} \ ; \ b = C \begin{pmatrix} \mathbb{1}_m \\ (0)_m \end{pmatrix}$$

For the initial point $v_0 = \lambda_0$, we will take $v_0 = \frac{1}{2} C \mathbb{1}$ $\leq C\mathbb{1}$

For the implementation, please see my Python code.