

# Convex optimization Exam

Raphael Reme

December 2020

## 1 Exercise 1

The problem  $(P)$  can be expressed as an LP problem:

$$(P) : \begin{array}{ll} \text{minimize}_x & \|x\|_\infty \\ \text{subject to} & Ax = b \end{array} \Leftrightarrow \begin{array}{ll} \text{minimize}_{x,z} & z \\ \text{subject to} & Ax = b \\ & z = \|x\|_\infty \end{array} \Leftrightarrow \begin{array}{ll} \text{minimize}_{x,z} & z \\ \text{subject to} & Ax = b \\ & z \geq |x_i|, \forall i \end{array} \quad (\text{Same as in HW2, Ex 3})$$

$$(P) \Leftrightarrow \begin{array}{ll} \text{minimize}_{x,z} & z \\ \text{subject to} & Ax = b \\ & \forall i, x_i - z \leq 0, \\ & -x_i - z \leq 0 \end{array} \Leftrightarrow \begin{array}{ll} \text{minimize}_{x,z} & z \\ \text{subject to} & Ax = b \\ & G \begin{pmatrix} x \\ z \end{pmatrix} \leq 0 \end{array}$$

With  $G = \begin{pmatrix} I_n & -1_n \\ -I_n & -1_n \end{pmatrix}$ . The lagrangian is therefore:

$$\begin{aligned} \mathcal{L}(x, z, \lambda, \nu) &= z + \lambda^T G \begin{pmatrix} x \\ z \end{pmatrix} + \nu^T (Ax - b) \\ &= z + \sum_{i=1}^n \lambda_i (x_i - z) + \sum_{i=n+1}^{2n} \lambda_i (-x_i - z) + \nu^T Ax - \nu^T b \\ &= z(1 - \lambda^T 1_{2n}) + \sum_{i=1}^n x_i (\lambda_i - \lambda_{n+i}) + \nu^T Ax - \nu^T b \\ &= z(1 - \lambda^T 1_{2n}) + (A^T \nu + \lambda_{1:n} - \lambda_{n+1:2n})^T x - b^T \nu \end{aligned}$$

We can now compute the dual function:  $g(\lambda, \nu) = \inf_{x,z} \mathcal{L}(x, z, \lambda, \nu)$ :

$$g(\lambda, \nu) = \begin{cases} -b^T \nu & \text{if } \lambda^T 1_{2n} = 1 \text{ and } A^T \nu + \lambda_{1:n} - \lambda_{n+1:2n} = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Thus the dual problem is:

$$(D) : \begin{array}{ll} \text{maximize}_{\lambda, \nu} & -b^T \nu \\ \text{subject to} & \lambda^T 1_{2n} = 1 \\ & A^T \nu + \lambda_{1:n} - \lambda_{n+1:2n} = 0 \\ & \lambda \geq 0 \end{array} \Leftrightarrow \begin{array}{ll} \text{maximize}_{\lambda_1, \lambda_2, \nu} & -b^T \nu \\ \text{subject to} & \lambda_1^T 1_n + \lambda_2^T 1_n = 1 \\ & A^T \nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \geq 0, \lambda_2 \geq 0 \end{array}$$

## 2 Exercise 2

Let  $u, v \in \mathbb{R}_{++}^n$ ,  $D_{KL}(u, v) = f(u) - f(v) - \nabla f(v)^T(u - v)$ , with  $f(x) = \sum_i x_i \log x_i$ .

Let's show that  $u \neq v \Rightarrow D_{KL}(u, v) > 0$ .

Let's consider the function  $h(x) = x \log x, \forall x \in \mathbb{R}_+^*$ ,  $h'(x) = 1 + \log x$  and  $h''(x) = \frac{1}{x} > 0$ . Therefore  $h$  is strictly

convex.

Now let's show that  $f$  is strictly convex: Let  $t \in ]0, 1[$ ,  $f(tu + (1-t)v) = \sum_i f(tu_i + (1-t)v_i) < \sum_i tf(u_i) + (1-t)f(v_i) = tf(u) + (1-t)f(v)$

(Note: it still holds for  $u = 0$  or  $v = 0$  with  $h(0) = 0$  and thus  $f(0) = 0$ )

Therefore as  $f$  is strictly convex and differentiable,  $\forall u \neq v, f(u) > f(v) + \nabla f(v)^T(u - v) \Rightarrow D_{KL}(u, v) > 0$ .

As  $D_{KL}(u, u) = f(u) - f(u) + 0 = 0$ ,  $\forall u, v \in \mathbb{R}^n, D_{KL}(u, v) \geq 0$  and  $D_{KL}(u, v) = 0 \Leftrightarrow u = v$ .

### 3 Exercise 3

The problem  $(P)$  can be simplified using that  $A = V^T V$  (and that the square root is a non decreasing function)

$$\begin{aligned} (P) : \text{minimize}_x \quad & c^T x & \Leftrightarrow \text{minimize}_x \quad & c^T x & \Leftrightarrow \text{minimize}_x \quad & c^T x \\ \text{subject to} \quad & x^T(A - bb^T)x \leq 0 & \text{subject to} \quad & (Vx)^T(Vx) \leq (b^T x)^2 & \text{subject to} \quad & \|Vx\|_2 \leq b^T x \\ & b^T x \geq 0 & & b^T x \geq 0 & & Dx = g \\ & Dx = g & & Dx = g & & \end{aligned}$$

Therefore  $(P)$  is a SOCP problem and thus a convex problem.

The dual can then be retrieve by introducing new variables:

$$\begin{aligned} (P) \Leftrightarrow \text{minimize}_{x,y,t} \quad & c^T x \\ \text{subject to} \quad & \|y\|_2 \leq t \\ & Dx = g \\ & y = Vx \\ & t = b^T x \end{aligned}$$

The lagrangian is

$$\begin{aligned} \mathcal{L}(x, y, t, \lambda, \nu_1, \nu_2, \nu_3) &= c^T x + \lambda(\|y\|_2 - t) + \nu_1^T(Dx - g) + \nu_2^T(y - Vx) + \nu_3(t - b^T x) \\ &= (c + D^T \nu_1 - V^T \nu_2 - \nu_3 b)^T x + \lambda\|y\|_2 + \nu_2^T y + (\nu_3 - \lambda)t - \nu_1^T g \end{aligned}$$

The dual function is

$$\begin{aligned} g(\lambda, \nu_1, \nu_2, \nu_3) &= \inf_{x,y,t} \mathcal{L}(x, y, t, \lambda, \nu_1, \nu_2, \nu_3) \\ &= \inf_x (c + D^T \nu_1 - V^T \nu_2 - \nu_3 b)^T x + \inf_y (\lambda\|y\|_2 + \nu_2^T y) + \inf_t t(\nu_3 - \lambda) - \nu_1^T g \\ &= \begin{cases} -\nu_1^T g & \text{if } \nu_3 = \lambda \text{ and } c + D^T \nu_1 - V^T \nu_2 - \nu_3 b = 0 \text{ and } \|\nu_2\| \leq \lambda \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

Using the fact that  $\inf_y (\lambda\|y\|_2 + \nu_2^T y) = \begin{cases} 0 & \text{if } \|\nu_2\| \leq \lambda \\ -\infty & \text{otherwise} \end{cases}$  (Can be shown with Cauchy-Schwarz inequality.

It has been done in previous homework) The dual problem is therefore:

$$\begin{aligned} (D) : \text{maximize}_{\lambda, \nu_1, \nu_2, \nu_3} \quad & -\nu_1^T g & \Leftrightarrow \text{maximize}_{\lambda, \nu_1, \nu_2} \quad & -g^T \nu_1 \\ \text{subject to} \quad & \|\nu_2\| \leq \lambda & \text{subject to} \quad & \|\nu_2\| \leq \lambda \\ & c + D^T \nu_1 - V^T \nu_2 - \nu_3 b = 0 & & c + D^T \nu_1 - V^T \nu_2 - \lambda b = 0 \\ & \nu_3 = \lambda & & \\ & \lambda \geq 0 & & \end{aligned}$$

## 4 Exercise 4

Let's compute the dual of  $(P)$ :

$$(P) : \underset{x}{\text{minimize}} \quad - \sum_i \log(b_i - a_i^T x) \Leftrightarrow \underset{x,y}{\text{minimize}} \quad - \sum_i \log(y_i) \\ \text{subject to} \quad y = b - Ax$$

(With  $A = \begin{pmatrix} a_1^T \\ \vdots \\ a_n^T \end{pmatrix}$ ). Which give the lagrangian:

$$\begin{aligned} \mathcal{L}(x, y, \nu) &= - \sum_i \log(y_i) + \nu^T (y - b + Ax) \\ &= -\nu^T b + \nu^T Ax + \sum_i \nu_i y_i - \log(y_i) \end{aligned}$$

$$\begin{aligned} g(\nu) &= \inf_{x,y} \mathcal{L}(x, y, \nu) \\ &= -\nu^T b + \inf_x \nu^T Ax + \sum_i \inf_{y_i > 0} (\nu_i y_i - \log(y_i)) \end{aligned}$$

The first infimum over  $x$  is bounded if and only if  $A^T \nu = 0$  (and then is equal to 0). And the infimums over  $y_i$  are unbounded for  $\nu_i \leq 0$  ( $\nu_i y_i - \log(y_i) \xrightarrow{y_i \rightarrow +\infty} -\infty$ ) and for  $\nu_i > 0$  the infimums are realised with  $y_i = \frac{1}{\nu_i}$  (From derivative)

$$\text{Thus } g(\nu) = \begin{cases} -\nu^T b + \sum_i 1 + \log(\nu_i) & \text{if } A^T \nu = 0 \text{ and } \nu > 0 \\ -\infty & \text{otherwise} \end{cases}$$

And the dual problem is:

$$(D) : \underset{\nu}{\text{maximize}} \quad -\nu^T b + m + \sum_i \log(\nu_i) \\ \text{subject to} \quad A^T \nu = 0$$

## 5 Exercise 5

Let the problems  $(P)$  and  $(\hat{P})$ :

$$(P) : \underset{x}{\text{minimize}} \quad f_0(x) \quad \text{and} \quad (\hat{P}) : \underset{x}{\text{minimize}} \quad f_0(x) + \alpha \|Ax - b\|_2^2 \\ \text{subject to} \quad Ax = b$$

The dual function of  $(P)$  is

$$\begin{aligned} g(\nu) &= \inf_x \mathcal{L}(x, \nu) \\ &= \inf_x (f_0(x) + \nu^T (Ax - b)) \end{aligned}$$

Let's  $\hat{x}$  a minimizer of  $(\hat{P})$ . As  $f_0$  is differentiable and convex, this minimizer verify  $\nabla f_0(\hat{x}) + 2\alpha A^T (A\hat{x} - b) = 0$ . Let  $\nu$ , one can also see that  $x$  minimizes  $f_0(x) + \nu^T (Ax - b)$  if and only if  $\nabla f_0(x) + A^T \nu = 0$

Let's called  $\hat{\nu} = 2\alpha(A\hat{x} - b)$  then a minimizer of  $f_0(x) + \hat{\nu}^T (Ax - b)$  is  $\hat{x}$  as  $\nabla f_0(\hat{x}) + A^T \hat{\nu} = \nabla f_0(\hat{x}) + 2\alpha A^T (A\hat{x} - b) = 0$ .

Therefore  $g(\hat{\nu}) = f_0(\hat{x}) + \hat{\nu}^T (A\hat{x} - b) = f_0(\hat{x}) + 2\alpha \|A\hat{x} - b\|_2^2 \Rightarrow \hat{\nu}$  is a feasible point of the dual! And we have a lower bound for our problem:

$$\begin{aligned} p = f_0(x^*) &\geq d \\ &= g(\nu^*) \\ &\geq g(\hat{\nu}) \end{aligned}$$

Finally:  $f_0(x^*) \geq f_0(\hat{x}) + 2\alpha \|A\hat{x} - b\|_2^2$

## 6 Exercise 6

Let the problem (P) (The same as SVM from Homework 2)

$$(P) : \begin{aligned} & \text{minimize}_{w,z} \quad \frac{1}{2} \|w\|_2^2 + C1_m^T z \\ & \text{subject to} \quad y_i w_i^T x_i \geq 1 - z_i, \forall i \in \llbracket 1, m \rrbracket \\ & \quad \quad \quad z \geq 0 \end{aligned}$$

We can recompute the dual the same way as before: with first the lagrangian:

$$\begin{aligned} \mathcal{L}(w, z, \lambda, \pi) &= \frac{1}{2} \|w\|_2^2 + C1_m^T z + \sum_i \lambda_i (1 - y_i w_i^T x_i - z_i) - \pi^T z \\ &= 1_m^T \lambda + (C1_m - \pi - \lambda)^T z + \frac{1}{2} \left( \|w\|_2^2 - 2 \left( \sum_i \lambda_i y_i x_i \right)^T w \right) \end{aligned}$$

As  $\inf_w \|w\|_2^2 - 2t^T w = -\|t\|_2^2$  (Using derivative). Then the dual problem is:

$$(D) : \begin{aligned} & \text{maximize}_{\lambda, \pi} \quad 1_m^T \lambda - \frac{1}{2} \left\| \sum_i \lambda_i y_i x_i \right\|_2^2 \Leftrightarrow \text{maximize}_{\lambda} \quad 1_m^T \lambda - \frac{1}{2} \left\| \sum_i \lambda_i y_i x_i \right\|_2^2 \\ & \text{subject to} \quad C1_m = \pi + \lambda \quad \text{subject to} \quad 0 \leq \lambda \leq C1_m \\ & \quad \quad \quad \lambda \geq 0, \pi \geq 0 \end{aligned}$$

(D) Can be rewritten as a QP program:

$$\begin{aligned} \left\| \sum_i \lambda_i y_i x_i \right\|_2^2 &= \left( \sum_i \lambda_i y_i x_i \right)^T \sum_i \lambda_i y_i x_i \\ &= \sum_{i,j} \lambda_i y_i x_i^T x_j y_j \lambda_j \\ &= \sum_{i,j} \lambda_i y_i (X X^T)_{i,j} y_j \lambda_j \quad (\text{With } X = \begin{pmatrix} x_1^T \\ \vdots \\ x_m^T \end{pmatrix} \in \mathbb{R}^{m \times n}) \\ &= \lambda^T \text{Diag}(y) X X^T \text{Diag}(y) \lambda \end{aligned}$$

And then

$$(D) \Leftrightarrow \begin{aligned} & \text{minimize}_{\lambda} \quad \frac{1}{2} \lambda^T \text{Diag}(y) X X^T \text{Diag}(y) \lambda - 1_m^T \lambda \Leftrightarrow \text{minimize}_{\lambda} \quad \lambda^T Q \lambda + p^T \lambda \\ & \text{subject to} \quad 0 \leq \lambda \leq C1_m \quad \text{subject to} \quad A \lambda \leq b \end{aligned}$$

With  $Q = \frac{1}{2} \text{Diag}(y) X X^T \text{Diag}(y)$ ,  $p = -1_m$ ,  $A = \begin{pmatrix} I_m \\ -I_m \end{pmatrix}$  and  $b = \begin{pmatrix} C1_m \\ 0_m \end{pmatrix}$ . Note that  $Q \geq 0$ : indeed by construction of  $Q$  we have  $\forall z, z^T Q z = \|\sum_i z_i y_i x_i\|_2^2 \geq 0$

We can now use the barrier method to solve the dual! (See the code) As the primal is strictly feasible and convex, strong duality holds (Slater's condition) and therefore the KKT conditions holds.

Let's call  $w^*, z^*, \lambda^*, \pi^*$  the solution of the primal/dual:

$$\begin{aligned} \text{Primal feasibility:} & \quad z^* \geq 0 \text{ and } y_i w_i^{*T} x_i \geq 1 - z_i^*, \forall i \in \llbracket 1, m \rrbracket \\ \text{Dual feasibility:} & \quad \lambda^* \geq 0, \pi^* \geq 0 \text{ and } C1_m = \pi^* + \lambda^* \\ \text{Complementary slackness:} & \quad \forall i \in \llbracket 1, m \rrbracket \pi_i^* z_i^* = 0, \lambda_i^* (1 - z_i^* - y_i w_i^{*T} x_i) = 0 \\ \text{First order conditions:} & \quad \nabla_w \mathcal{L}(w^*, z^*, \lambda^*, \pi^*) = 0, \nabla_z \mathcal{L}(w^*, z^*, \lambda^*, \pi^*) = 0 \end{aligned}$$

We can retrieve  $w^*, z^*$  ( $\pi^*$ ) from  $\lambda^*$  (obtained with the barrier method):

$$w^* = \sum_i \lambda_i^* y_i x_i = X^T \text{Diag}(y) \lambda^* \quad (\text{first order condition on } w)$$

$$\pi^* = C1_m - \lambda^* \quad (\text{Dual feasibility})$$

$$z_i^* = \begin{cases} 0 & \text{if } \pi_i^* \neq 0 \Leftrightarrow \lambda^* < C \\ 1 - y_i w^{*T} x_i & \text{if } \lambda_i^* > 0 \end{cases} \quad (\text{Complementary slackness})$$

Please see the provided notebook to see the code.