CONVEX OPTIMIZATION HOMEWORK 4

TOMMENIX YU
ID: 12370130
STAT 31015
DUE WED FEB 15, 2023, 3PM

Exercise 1.

Proof.

First, we decompose $x := \begin{pmatrix} y \\ z \end{pmatrix} := (y, z)$ for $y \in \mathbb{R}^p$, $z \in \mathbb{R}^{n-p}$.

Now, define

$$H(y,z) = \begin{pmatrix} h_1(y,z) \\ \vdots \\ h_n(y,z) \end{pmatrix}$$

and we get

$$DH(y,z) = \left(\begin{array}{ccc|ccc} \frac{\partial h_1}{\partial y_1}(y,z) & \dots & \frac{\partial h_1}{\partial y_p}(y,z) & \frac{\partial h_1}{\partial z_1}(y,z) & \dots & \frac{\partial h_1}{\partial z_{n-p}}(y,z) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial h_p}{\partial y_1}(y,z) & \dots & \frac{\partial h_p}{\partial y_p}(y,z) & \frac{\partial h_p}{\partial z_1}(y,z) & \dots & \frac{\partial h_p}{\partial z_{n-p}}(y,z) \end{array}\right) =: [A|B].$$

We know that at $x^* = (y^*, z^*)$ the Jacobi is full row rank so we can rearrange the Jacobi with the first p columns being the pivot columns in the row reduced echelon form. Since this can be done with a permutation in terms of x, we just assume WLOG that the first p columns are the pivot rows to avoid re-defining everything.

Now we know that the matrix A is full rank since it has all pivot columns. In particular it is invertible, and thus we can apply the inverse function theorem to get that $\exists! g : \mathbb{R}^{n-p} \to \mathbb{R}^p$ unique continuously differentiable such that $H(g(z), z) = 0$ in a neighborhood of z^*, since x^* satisfies the problem, so $H(x^*) = 0$.

So now the problem becomes an unconstrained problem

$$\min \quad F(z)$$

where $F(z) := f(g(z), z)$.

Thus, for x^* the minimum of the new problem, it must have that $F'(z^*) = 0$ where $x^* = (g(z^*), z^*)$. Taking the derivative of the function we get

$$G(z) := (g(z), z)$$

to get

$$\nabla F^T = \nabla f(g(z),z)^T \nabla G(z) = \begin{pmatrix} Dg \\ I_{n-p} \end{pmatrix}^T_{n\times(n-p)} \begin{pmatrix} \partial_1 f(g(z),z) \\ \vdots \\ \partial_n f(g(z),z) \end{pmatrix}$$

and by chain rule with some rearrangements (IFT gives $Dg = -A^{-1}B$) we get

$$0 = \nabla F = (-A^{-1}B)^T \nabla_y f + \nabla_z f$$

which we write as

$$\nabla_z f = B^T (A^{-1})^T \nabla_y f$$

and note that

$$\nabla_y f = A^T (A^{-1})^T \nabla_y f$$

combining we get

$$\nabla f = \begin{pmatrix} \nabla_y f \\ \nabla_z f \end{pmatrix} = \begin{pmatrix} A^T \\ B^T \end{pmatrix} (A^{-1})^T \nabla_y f = [A|B]^T (A^{-1})^T \nabla_y f = DH^T (A^{-1})^T \nabla_y f$$

$$\Rightarrow \nabla f - DH^T (A^{-1})^T \nabla_y f = 0$$

where we just let $\nu = -(A^{-1})^T \nabla_y f$ to get

$$\nabla f + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

and since our A is fixed, $\nabla_y f$ is fixed, and the choice of this relation is unique by IFT, the choice of ν is unique.

$\square$

Exercise 2.

Proof.

Step 1: simplify the problem to only the active conditions

We prove that if KKT holds for only the active conditions, then it is also a solution to the original problem.

That is, if we know

$$\nabla f(x^*) + \sum_{m'}^{i=1} \mu_i \nabla g_i(x^*) + \sum_{p}^{i=1} \nu_i \nabla h_i(x^*) = 0$$

then by adding 0 times all the inactive conditions the gradient condition also holds. Moreover, adding zeros does not affect the ≥ 0 condition for μ.

Step 2: Switch to equality conditions

Since all that's left is active inequality conditions and equality constraints, we can pack them together to be a equality constraint question. Thus, we apply question 1 to get the existence result of such μ and ν with

$$\nabla f(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{i=1}^{p} \nu_i \nabla h_i(x^*) = 0.$$

Step 3: feasibily of dual conditions

We need to show $\mu \geq 0$, where μ is as in step 2. For contradiction, assume that there is an $\mu_i < 0$. That is, i is fixed here. Moreover, let's define our new

$$H(x) = \left(h_1(x), \dots, h_p(x), g_1(x), \dots, g_{i-1}(x), g_{i+1}(x), \dots, g_m(x), \right)^T : \mathbb{R}^n \to \mathbb{R}^{m+p-1}$$

and here we need to assume $n \geq m + p$ to use the implicit function theorem, in the exactly same manner (module change of variables, and $m + p - 1$ in place of m) as in problem 1 to rearrange pivot columns of DH and rewrite $x = (y, z)$ for $y \in \mathbb{R}^{m+p-1}$, $z \in \mathbb{R}^{n-m-p+1}$ to attain (since linear independent) a unique C^1 $\phi : \mathbb{R}^{n-m-p+1} \to \mathbb{R}^{m+p-1}$ such that

$$H(\phi(z), z) = 0$$

for z near z^*.

Now, since $g_i'(x^*)$ is linearly independent with other vectors it is not 0, so we know that there is a basis direction k in which $(g_i'(x^*))_k \neq 0$. Moreover, since it is linearly independent with all $h_j'(x^*)$ and $g_j'(x^*)$ for $j \neq i$, we know that at least one of such basis direction k is not the pivot columns of DH. Since we can WLOG rearrange everything as in prob 1, let's assume that the $(m + p)$th term of the derivative $g_i'(x^*)$ is non-zero.

This gives us the existence of a direction $v \in \mathbb{R}^{n-m-p+1}$ with $g_i(y^*, z^* + v) = -1$ (which in most cases is far from unique), and let's fix this direction and call it v.

Now let's rewrite

$$x = x(t) =: (\phi(z^* + c(t)v), z^* + c(t)v)$$

with $c(t)$ satisfying $g_i(x(t)) = -t$ in the neighborhood.

By step 2 we get

$$\nabla f(x(t)) = -\sum_{j=1}^{m} \mu_j \nabla g_j(x(t)) - \sum_{j=1}^{p} \nu_j \nabla h_j(x(t))$$

and because of our construction of $x(t)$ is based on ϕ, which makes $H(x(t)) = 0$ along the direction, the right hand side is left only with the term

$$\frac{d}{dt} f(x(t)) = -\mu_i \frac{d}{dt} g_i(x(t)) = (-\mu_i) \cdot \left(c \cdot g_i(y^*, z^* + v)\right) < 0$$

for some constant c by construction above.

But this cannot be since $x^* = x(0)$ is the minimal point of f, so for any direction the directional derivative should be ≥ 0. Hence contradiction! So we cannot have $\mu_i < 0$.

Step 4: state the KKT condition

(1) (Primal constraints) $g_i(x) \leq 0$, $h_i(x) = 0$.
(2) (dual constraints) $\mu \succeq 0$ (by step 1,2,3).
(3) (complementary slackness) $\mu_i g_i = 0$ (by step 1 + definition of active conditions).
(4) (Gradient = 0)

$$\nabla f(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{i=1}^{p} \nu_i \nabla h_i(x^*) = 0$$

by step 2 and problem 1.

□

Exercise 3. *Source:* https://angms.science/doc/CVX/Proj_l1.pdf

Proof.

The Lagrangian is

$$L(x, \lambda) = \frac{1}{2}||x - x_0||^2 + \lambda(1 \cdot |x| - 1)$$

for which we want to find the minimum over x, so take the derivative to 0 and get

$$x - x_0 + \lambda \operatorname{sgn}(x) = 0$$

where $\operatorname{sgn}(x) = (\operatorname{sgn}(x_1), \ldots, \operatorname{sgn}(x_n))^T$.

Since the domain is a closed unit ball under the L^1 norm, it has non-empty interior, which means that KKT conditions hold.

So we list

(1) $||x||_1 \le 1$.
(2) $\lambda \ge 0$.
(3) $\lambda||x||_1 - \lambda = 0$.
(4) $x - x_0 + \lambda \operatorname{sgn}(x) = 0$.

If $\lambda = 0$, then we get $x = x_0$ and the minimum is 0.

If $\lambda > 0$, then $||x||_1 - 1 = 0$, so x must be on the L^1 unit ball (the unit diamond). But here we can further written as

$$x_i = \operatorname{sgn}(x_{0,i}) \max\{|x_{0,i}| - \lambda, 0\}$$

because we can separate and deal for all 4 cases:

(1) $x_i \ge 0$, $x_{0,i} \ge 0$: $\operatorname{sgn}(x_{0,i}) \max\{|x_{0,i}| - \lambda, 0\} = x_{0,i} - \lambda = x_0 \ge 0$;
(2) $x_i < 0$, $x_{0,i} \ge 0$: $\Rightarrow x_{0,i} + \lambda < 0 \Rightarrow \lambda < 0$ contradiction to $\lambda > 0$, so trivially true;
(3) $x_i \ge 0$, $x_{0,i} < 0$: since $x_i = x_{0,i} - \lambda$ here so $\lambda < 0$, contradiction. Again this case is trivially true;
(4) $x_i < 0$, $x_{0,i} < 0$: $\operatorname{sgn}(x_{0,i}) \max\{|x_{0,i}| - \lambda, 0\} = -(-x_{0,1} - \lambda) = x_{0,1} + \lambda = x_0$.

Thus, we plug in and get

$$\sum_{i=1}^{n} |x_i| = 1 \Rightarrow \sum_{i=1}^{n} |\operatorname{sgn}(x_{0,i}) \max\{|x_{0,i}| - \lambda, 0\}| = 1$$

$$\Rightarrow \sum_{i=1}^{n} \max\{|x_{0,i}| - \lambda, 0\} = 1$$

Again, we separate cases:

(1) $x_{0,i} > \lambda > 0$: $x_i = x_{0,i} - \lambda$;
(2) $-\lambda < x_{0,i} < \lambda$: $x_i = 0$;
(3) $x_{0,i} < -\lambda$: $x_i = \lambda + x_{0,i}$

and so we've found that the formula for the optimal x:

$$x = f_\lambda(x_0) = d(x_0, [-\lambda, \lambda]) = \begin{cases} x_{0,i} - \lambda & x_{0,i} > \lambda > 0 \\ 0 & -\lambda < x_{0,i} < \lambda \\ x_{0,i} + \lambda & x_{0,i} < -\lambda \end{cases}$$

This is related to the dual problem because if we compute the dual we get

So the dual problem

$$g(\lambda) = \sum_{i=1}^{n} \left[\frac{1}{2}(\operatorname{sgn}(x_{0,i}) \max\{|x_{0,i}| - \lambda, 0\}) - x_{0,i})^2 + \lambda \max\{|x_{0,i}| - \lambda, 0\} \right] - \lambda$$

where we again separate case to get via computation

$$\frac{1}{2}(\operatorname{sgn}(x_{0,i}) \max\{|x_{0,i}| - \lambda, 0\}) - x_{0,i})^2 + \lambda \max\{|x_{0,i}| - \lambda, 0\} = \begin{cases} -\frac{1}{2}\lambda^2 + \lambda x_{0,i} & x_{0,i} > \lambda \\ \frac{1}{2}x_{0,i}^2 & -\lambda < x_{0,i} < \lambda \\ -\frac{1}{2}\lambda^2 - x_{0,i}\lambda & x_{0,i} < -\lambda \end{cases}$$

where both the first case and the third case's maximum is attained when $\lambda = |x_{0,i}|$ i.e. at the smalles endpoint of that interval.

Thus, we are trying to maximize

$$\sum_{i=1}^{n} h(n) - \lambda$$

where h is the function computed above. Using the expression to get a maximum point of the function λ^*. So the only we need to do is to plug back into the optimal x formula and get

$$x_j = \begin{cases} x_{0,i} - \lambda^* & x_{0,i} > \lambda^* > 0 \\ 0 & -\lambda^* < x_{0,i} < \lambda^* \\ x_{0,i} + \lambda^* & x_{0,i} < -\lambda^* \end{cases}$$

□

Exercise 4.

Proof.

$\phi(x)$ is convex:

Since x^2 is convex and non-decreasing on $[0, \infty)$ and $\max 0, f_i$ is the maximum of convex functions, so it is convex. Moreover, it's range is $[0, \infty)$. Thus, by combination theorem in class, $x^2 \circ \max\{0, f_i(x)\} = \max\{0, f_i\}^2$ is still convex. Since $\alpha > 0$ and f_0 convex, the summation of convex functions is convex, so is ϕ.

Feasible point and lower bound:

Since $\tilde{x}$ minimizes problem, we have that

$$\nabla\phi = \nabla f + 2\alpha \sum_{i=1}^{n} \max\{0, f_i\} \nabla f.$$

For the original problem, the Lagrangian is

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

and taking derivative to 0 we get that all

$$\nabla L(x, \lambda) = \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f(x) = 0.$$

But really we plug in $\tilde{x}$ and note that if $\lambda_i = 2\alpha \max\{0, f_i(\tilde{x})\}$ we have

$$\nabla L(\tilde{x}, \lambda) = 0$$

and hence

$$g(\lambda) = \inf_x L(x, \lambda)$$

which since there is a dual solution, there is a feasible point.

Moreover

$$p^* \geq d^* = \sup_{\mu \geq 0} g(\mu) \geq g(\lambda) = f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{x})$$

is thus a lower bound.

□

Exercise 5.

Proof.

(1): We've learned in class that the two problems in question are duals of each other so we just write one code and use the dual variable result to get the solution for the second problem.

Code is called "q5runcode" for this prob.

The result for $\varepsilon = 0$ is

```
x =

    1.0000
    1.0000
    1.0000
    1.0000


y =

    0.0000
    2.0000
    1.0000
    2.0000
    2.0000
```

The result between $[-0.1, 0.1]$ is the following.

```
H =

  列 1 至 13

   -0.1000   -0.0900   -0.0800   -0.0700   -0.0600   -0.0500   -0.0400   -0.0300   -0.0200   -0.0100         0    0.0100    0.0200
    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000
    0.9000    0.9100    0.9200    0.9300    0.9400    0.9500    0.9600    0.9700    0.9800    0.9900    1.0000    1.0100    1.0200
    0.8000    0.8200    0.8400    0.8600    0.8800    0.9000    0.9200    0.9400    0.9600    0.9800    1.0000    1.0200    1.0400
    1.1000    1.0900    1.0800    1.0700    1.0600    1.0500    1.0400    1.0300    1.0200    1.0100    1.0000    0.9900    0.9800

  列 14 至 21

    0.0300    0.0400    0.0500    0.0600    0.0700    0.0800    0.0900    0.1000
    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000
    1.0300    1.0400    1.0500    1.0600    1.0700    1.0800    1.0900    1.1000
    1.0600    1.0800    1.1000    1.1200    1.1400    1.1600    1.1800    1.2000
    0.9700    0.9600    0.9500    0.9400    0.9300    0.9200    0.9100    0.9000


F =

  列 1 至 13

   -0.1000   -0.0900   -0.0800   -0.0700   -0.0600   -0.0500   -0.0400   -0.0300   -0.0200   -0.0100         0    0.0100    0.0200
    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000
    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000
    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000
    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000
    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000

  列 14 至 21

    0.0300    0.0400    0.0500    0.0600    0.0700    0.0800    0.0900    0.1000
    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000
    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000
    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000
    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000
    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000    2.0000
```

(2): The dual problem's inequality constraint can be modified into

$$Ax - b \preceq \varepsilon f = u$$

as the notation in chapter 5.6. So $\dfrac{\partial u}{\partial \varepsilon} = f$.

Moreover, by multivariable calculus we get

$$\frac{\partial p^*(\varepsilon)}{\varepsilon} = \sum_{i=1}^{4} \frac{\partial p^*(\varepsilon)}{\partial u_i}\frac{\partial u_i}{\partial \varepsilon} = -\sum_{i=1}^{4} \lambda_i^* \cdot f_i = -\lambda^* \cdot f$$

using formula 5.58.

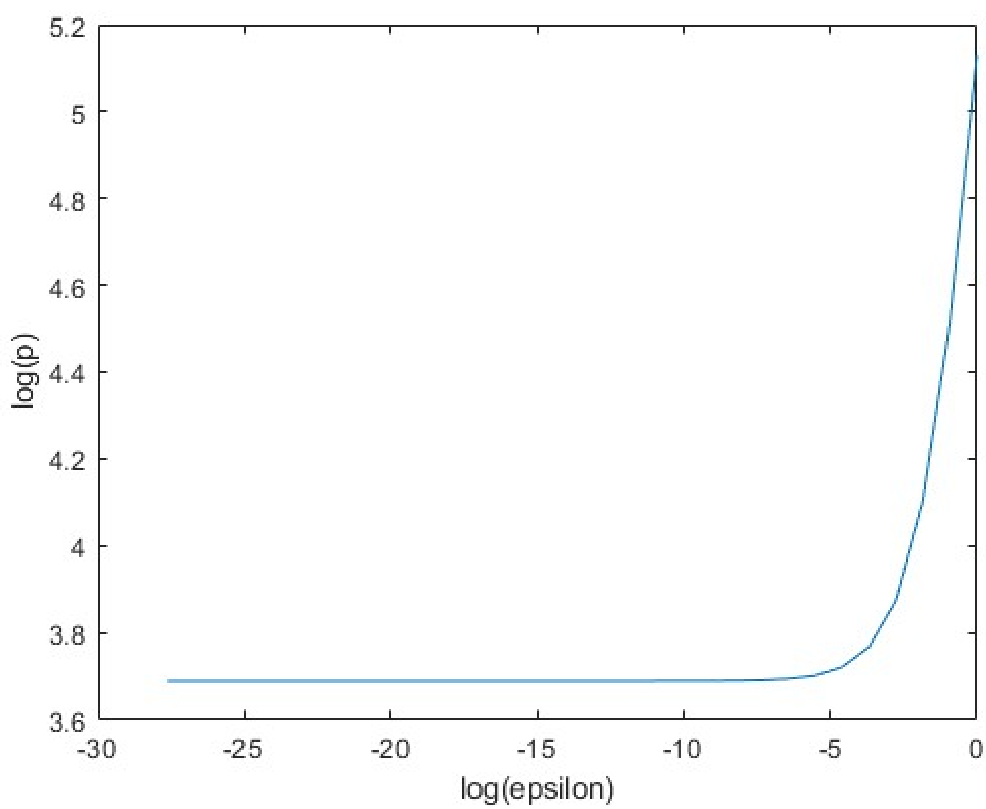
But then λ^* is nothing but the optimal point of the dual problem, i.e. the y we got with cvx. So we can just compute $-y^T f$ to get the result. The result we get for y is $(0, 2, 1, 2, 2)$ and hence matlab says

$$-y^T f = -129.0000$$

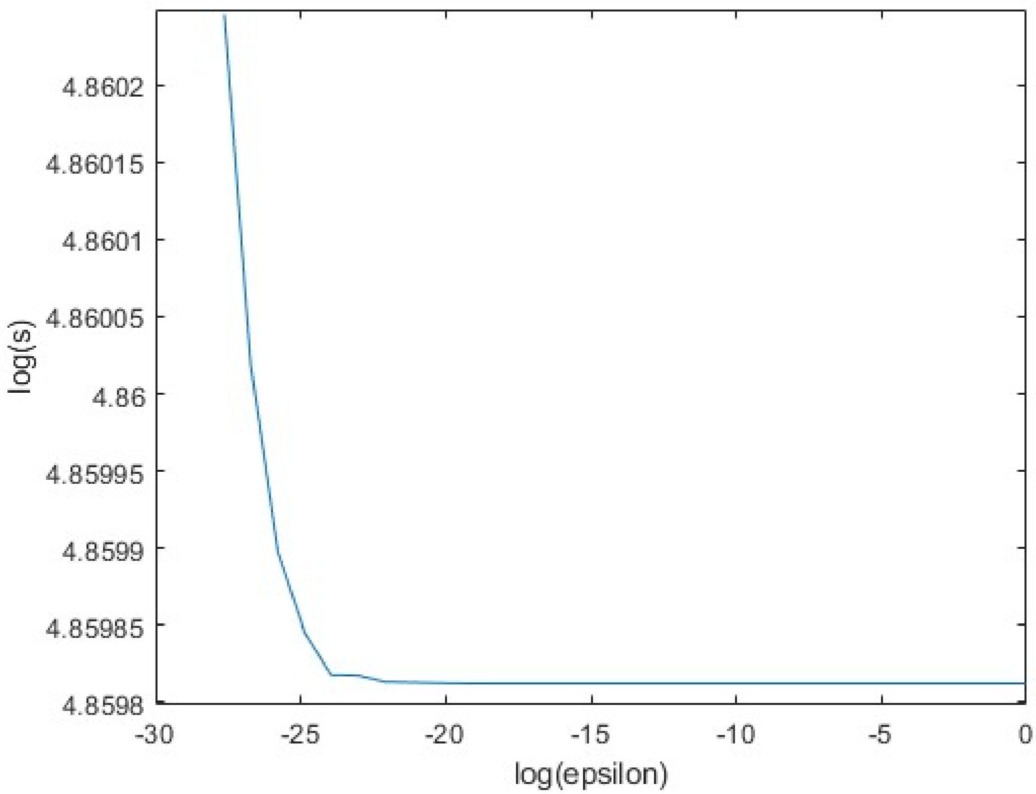
should be the answer.

(3): What we do is to use $\varepsilon = \delta$ small and compute p^*, which is the optimal value returned by cvx. The graph is the following:

For numerical value:



For sensitivity:



□