CME 307/MS&E 311: Optimization
Instructor: Prof. Yinyu Ye
Midterm Exam: Winter 2021-2022

HONOR CODE

In taking this examination, I acknowledge and accept Stanford University Honor Code.

NAME (Signed) : ______________________________

NAME (Printed): ______________________________

Problem	Full Points	Score
1	30	
2	20	
3	20	
4	20	
5	20	
Bonus	5	
Total	110 + 5	

CME 307/MS&E 311 | **Midterm**
Optimization | **Winter 2021-2022**
Prof.Yinyu Ye | **24 hours**

Note: You have 24 hours to work on the exam. Late submissions are not accepted. Upload your solutions back to Gradescope. In taking this examination, you acknowledge and accept the Stanford University Honor Code.

Question 1 (30') [True/False] Give a *true* or *false* answer to each of the following questions and explain your choice. (If your answer is *true*, provide an argument or cite the appropriate claims from lecture notes or textbook. If your answer is *false*, provide a counterexample or cite the appropriate claims from lecture notes or textbook).

(a) (5') In classical linear programming, even if the problem is feasible and bounded, the optimal solution might not be attained.

Solution False. See page 5 in Lecture Note 4. If the problem is feasible and bounded, its dual is also feasible and bounded. Thus, both primal and dual problems have optimal solutions.

(b) (5') In classical linear programming, strong duality always holds if both of the primal and the dual problems are feasible. In contrast, in semi-definite programming, there are feasible problem pairs where strong duality does not hold.

Solution True. For a couter-example, please see page 12 in Lecture Note 4.

(c) (5') In a convex constrained optimization problem, if $\mathbf{x}$ is a local minimizer, then it is a KKT solution.

Solution False. For a counter-example, see page 12 in Lecture Note 7. Any local minimizer is a KKT point if it is a regular point (Page 10 of Lecture Note 7) even if it is a convex optimization problem.

(d) (5') Consider a set C defined by $C := \{\mathbf{x} = (x_1, ..., x_n)^T \in \mathbb{R}^n_+ \, : \, f(\mathbf{x}) \geq 0\}$, where $n \geq 2$. If $f(\mathbf{x}) = \Pi_{i=1}^n x_i - 1$, then C is a convex set even $f(\mathbf{x})$ is a non-concave function on C.

Solution True. To verify C is a convex set, we write it in an equivalent form

$$C = \{\mathbf{x} : \sum_{i=1}^{n} -\log(x_i) \leq 0\},$$

which implies C is convex. To see $f(\mathbf{x})$ is not a concave function, directly check the Hessian matrix when $n = 2$. In this case, $f(\mathbf{x})$ is neither convex nor concave.

(e) (5') Consider a conic LP in the standard form

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c} \cdot \mathbf{x} \\ \text{s.t.} \quad & \boldsymbol{a}_i \cdot \mathbf{x} = b_i, \ i = 1, ..., m, \ (\mathcal{A}\mathbf{x} = \mathbf{b}) \\ & \mathbf{x} \in K. \end{aligned}$$

Similar to the Lagrangian function, we can construct

$$L(\mathbf{x}, \mathbf{y}, \mathbf{s}) = \mathbf{c} \cdot \mathbf{x} - \mathbf{y}^T(\mathcal{A}\mathbf{x} - \mathbf{b}) - \mathbf{s} \cdot \mathbf{x},$$

where $\mathbf{y} \in \mathbb{R}^m$ and $\mathbf{s} \in K^*$. Let

$$\phi(\mathbf{y}, \mathbf{s}) = \inf_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}, \mathbf{s}).$$

Then, the conic dual problem is equivalent to the lagrangian dual problem, that is, $\max_{\mathbf{y}, \mathbf{s} \in K^*} \phi(\mathbf{y}, \mathbf{s})$.

Solution True. If $\mathbf{c} - \mathbf{s} - \mathcal{A}^T\mathbf{y} \neq 0$, we have $\phi(\mathbf{y}, \mathbf{s}) = -\infty$ by taking $\mathbf{x} = \alpha(\mathbf{c} - \mathbf{s} - \mathcal{A}^T\mathbf{y})$ and letting $\alpha \to -\infty$. Otherwise, we have

$$L(\mathbf{x}, \mathbf{y}, \mathbf{s}) = \mathbf{b}^T\mathbf{y}.$$

Thus, the Lagrangian dual problem is eqivalent to

$$\max_{\mathbf{c}=\mathbf{s}+\mathcal{A}^T\mathbf{y}, \mathbf{s}\in K^*} \mathbf{b}^T\mathbf{y},$$

which is the conic dual problem.

(f) (5') Consider a feasible optimization problem on $\mathbb{R}^3$,

$$\begin{aligned} \min \quad & f(x_1, x_2, x_3) \\ \text{s.t.} \quad & c(x_1, x_2, x_3) \leq 0, \end{aligned}$$

where $f(\mathbf{x})$ and $c_i(\mathbf{x})$ are strongly convex functions.[1] Assume that $f(x_1, x_2, x_3) = f(x_1, x_3, x_2)$ and $c(x_1, x_2, x_3) = c(x_1, x_3, x_2)$, for any $x_1, x_2, x_3 \in \mathbb{R}$. Assume this problem has a minimizer. Then we must have $x_2 = x_3$ at the unique minimizer.

Solution TRUE. Otherwise we would have more than one minimizer, which is impossible because the function is strongly convex.

[1] A strongly convex function f satisfies $f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) < \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y})$ for distinct x, y and $\alpha \in (0, 1)$

Question 2 [SVM] (20')

Recall the Supporting Vector Machine in Lecture 1 and question 6 in Homework 2. Let the red class of points contain three points

$$\mathbf{a}_1 = (0,3), \mathbf{a}_2 = (1,2), \mathbf{a}_3 = (2,3),$$

and let the blue class of points contain three points

$$\mathbf{b}_1 = (1,0), \mathbf{b}_2 = (2,1), \mathbf{b}_3 = (1,3),$$

which are illustrated in Figure 1.

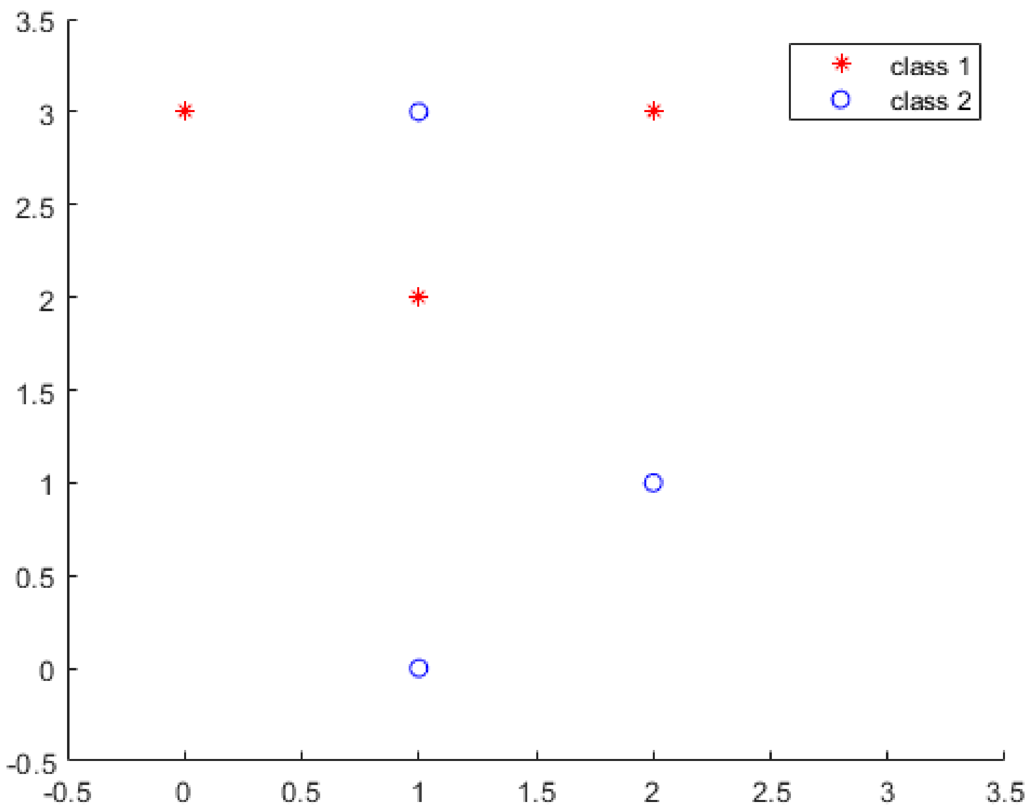

Figure 1: Scatter of all points

(a) (4') Consider the "hard-margin" SVM problem

$$\begin{aligned} \min \quad & \|\mathbf{x}\|^2 \\ \text{s.t.} \; & \mathbf{a}_i^T\mathbf{x} + x_0 \geq 1, i = 1,2,3 \\ & \mathbf{b}_j^T\mathbf{x} + x_0 \leq -1, j = 1,2,3. \end{aligned}$$

Is this problem feasible (i.e. there is a line that can separate the two classes of points)? Explain why.

Solution It is infeasible. $\mathbf{b}_1^T\mathbf{x} + x_0 \leq -1$ and $\mathbf{a}_2^T\mathbf{x} + x_0 \geq 1$ implies $\mathbf{x}(2) \geq 0$. However, it is contradict with $\mathbf{b}_3^T\mathbf{x} + x_0 \leq -1$.

(b) (6') Consider the following SVM problem with additional variable β (which was described in class):

$$\begin{aligned} \min \quad & \beta + \|\mathbf{x}\|^2 \\ \text{s.t.} \quad & \mathbf{a}_i^T\mathbf{x} + x_0 + \beta \geq 1, i = 1, 2, 3 \\ & \mathbf{b}_j^T\mathbf{x} + x_0 - \beta \leq -1, j = 1, 2, 3 \\ & \beta \geq 0. \end{aligned}$$

Show that this problem is feasible and find an optimal solution by hand. (Hint: you can try to show $\beta \geq 1$ at first, and then fix $\beta = 1$.)

Solution

Plugging in $\mathbf{a}_2 = (1, 2)$, $\mathbf{b}_1 = (1, 0)$, $\mathbf{b}_3 = (1, 3)$, we have

$$\begin{aligned} x_1 + 2x_2 + x_0 + \beta &\geq 1 \\ x_1 + x_0 - \beta &\leq -1 \\ x_1 + 3x_2 + x_0 - \beta &\leq -1 \end{aligned}$$

The first two inequalities imply that $2x_2 + 2\beta \geq 2$, while the first and the third inequalities imply that $-x_2 + 2\beta \geq 2$. It follows that $\beta \geq 1$ so that the optimal objective value is bounded below by 1. On the other hand, setting $\beta = 1$, $\mathbf{x} = 0$

and $x_0 = 0$ we see that this is a feasible solution. It follows that the optimal value is 1.

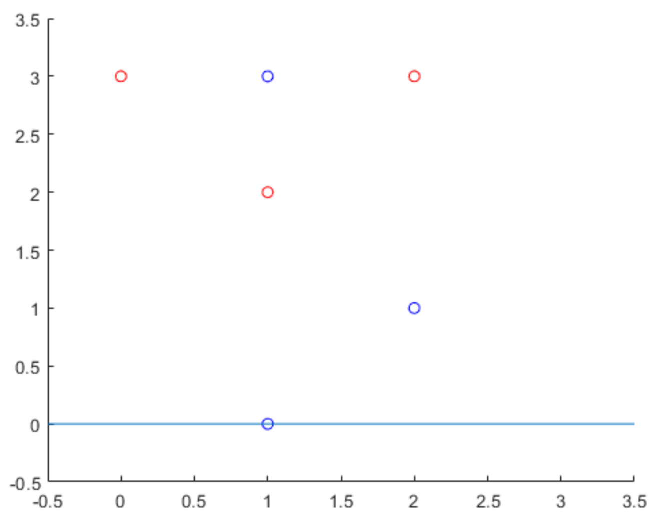

Figure 2: Scatter of all points and classifier 1

(c) (6') Consider the following "soft-margin" SVM problem with six additional variable β's:

$$\begin{aligned} \min \ & \frac{1}{6}(\beta_1^a + \beta_2^a + \beta_3^a + \beta_1^b + \beta_2^b + \beta_3^b) + \|\mathbf{x}\|^2 \\ \text{s.t. } & \mathbf{a}_i^T \mathbf{x} + x_0 + \beta_i^a \geq 1, i = 1,2,3 \\ & \mathbf{b}_j^T \mathbf{x} + x_0 - \beta_j^b \leq -1, j = 1,2,3 \\ & \beta_i^a, \beta_j^b \geq 0, \ i,j = 1,2,3. \end{aligned}$$

Please construct the dual of this "soft-margin" SVM problem. In this case, one optimal solution is

$$\begin{aligned} & \mathbf{x} = (-0.08, 0.33)^T, \ x_0 = -0.50, \\ & \beta_1^a = 0.51, \ \beta_2^a = 0.92, \ \beta_3^a = 0.67, \\ & \beta_1^a = 0.41, \ \beta_2^a = 0.66, \ \beta_3^b = 1.41. \end{aligned}$$

Solution With the Lagrangian function, we can have the dual problem is

$$\begin{aligned}
\max \quad & -\frac{1}{4}\|\sum_{i=1}^{3} y_i^a \boldsymbol{a}_i + \sum_{j=1}^{3} y_j^b \boldsymbol{b}_j\|^2 + \sum_{i=1}^{3} y_i^a - \sum_{j=1}^{3} y_j^b \\
\text{subject to} \quad & \sum_{i=1}^{3} y_i^a + \sum_{j=1}^{3} y_j^b = 0 \\
& \frac{1}{6} - y_i^a - s_i^a = 0, \text{ for } i = 1, 2, 3 \\
& \frac{1}{6} + y_j^b - s_j^b = 0, \text{ for } j = 1, 2, 3 \\
& y_i^a, s_i^a, s_j^b \geq 0 \ y_j^b \leq 0, \text{ for } i, j = 1, 2, 3
\end{aligned}$$

True numerical solution

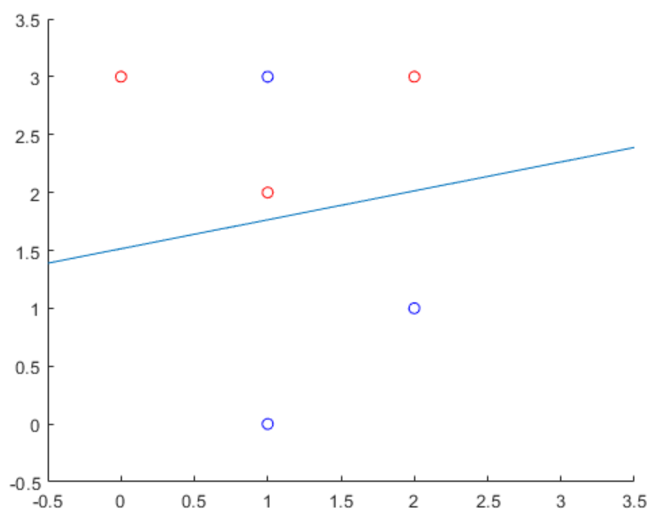

Figure 3: Scatter of all points and classifier - soft margins

Approximate solution version: The simplified soft-margin SVM is shown below:

$$\begin{aligned}
&\min\ \|\mathbf{x}\|^2 \\
&\text{s.t. } \mathbf{a}_1^T\mathbf{x} + x_0 \geq 0.49\ (\lambda_1^a), \\
&\quad\ \mathbf{a}_2^T\mathbf{x} + x_0 \geq 0.07\ (\lambda_2^a), \\
&\quad\ \mathbf{a}_3^T\mathbf{x} + x_0 \geq 0.32\ (\lambda_3^a), \\
&\quad\ \mathbf{b}_1^T\mathbf{x} + x_0 \leq -0.58\ (\lambda_1^b), \\
&\quad\ \mathbf{b}_2^T\mathbf{x} + x_0 \leq -0.33\ (\lambda_2^b), \\
&\quad\ \mathbf{b}_3^T\mathbf{x} + x_0 \leq 0.42\ (\lambda_3^b).
\end{aligned}$$

Then, we can find the dual problem:

$$\begin{aligned}
&\min\ 0.49\lambda_1^a + 0.07\lambda_2^a + 0.32\lambda_3^a + 0.58\lambda_1^b + 0.33\lambda_2^b - 0.42\lambda_3^b \\
&\text{s.t. } \sum_{i=1}^{3}\lambda_i^a - \sum_{j=1}^{3}\lambda_j^b = 0, \\
&\quad\ \lambda_i^a \geq 0,\ i = 1,2,3, \\
&\quad\ \lambda_j^a \geq 0,\ j = 1,2,3.
\end{aligned}$$

Then, based on the given conditions, we have

$$\begin{aligned}
\mathbf{a}_1^T\mathbf{x} + x_0 &= 0.49, \\
\mathbf{a}_3^T\mathbf{x} + x_0 &= 0.32, \\
\mathbf{b}_1^T\mathbf{x} + x_0 &= -0.58,
\end{aligned}$$

and can find the result that

$$(\mathbf{x}; x_0) = \begin{pmatrix} -1/2 & 1/2 & 0 \\ 1/6 & 1/6 & -1/3 \\ 1/2 & -1/2 & 1 \end{pmatrix} \begin{pmatrix} 0.49 \\ 0.32 \\ -0.58 \end{pmatrix}$$

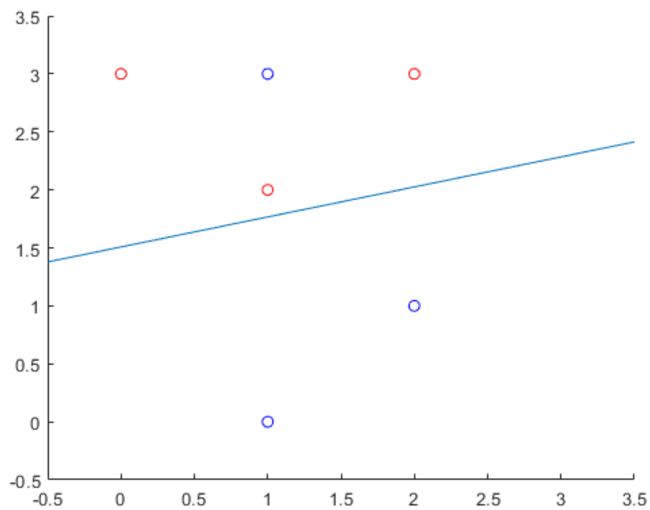

Figure 4: Scatter of all points and classifier - soft margin solved by hand

(d) (4') Compare the results of part (a), (b) and (c). Explain which model would you likely to choose to use for this case and why. What about in general? (There is no right or wrong answer to this question.)

Solution Comparing the two graphs above, I would choose the third one. Moreover, the third SVM is more robust to outliers. But any reasonable answer with SVMs in parts (a) and (b) is ok for this question.

Question 3 [Resource Allocation] (20')

A function $f : R^n \longrightarrow R$ is said to be *separable* if it can be written in the form

$$f(x_1, \dots, x_n) = \sum_{j=1}^{n} f_j(x_j).$$

Such functions were studied by Gibbs in connection with work on the chemical equilibrium problem (1876). He showed a theorem which states that (in the differentiable case) a necessary condition of local optimality for a feasible point x^* of the separable nonlinear programming problem

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}} & \displaystyle\sum_{j=1}^{n} f_j(x_j) \\ \text{subject to} & \displaystyle\sum_{j=1}^{n} x_j = M \quad (M > 0) \\ & x_j \geq 0 \quad j = 1, \dots, n \end{array}$$

is that there exist a thresholding number λ^* such that

$$f_j'(x_j^*) = \lambda^* \quad \text{if } x_j^* > 0$$

$$f_j'(x_j^*) \geq \lambda^* \quad \text{if } x_j^* = 0$$

Here f' is the derivative of f, and given M represents the total amount of the single resource. This constraint is called "Knapsack" constraint.

(a) (5') Show that every minimizer is a KKT solution. (Hint: at lease one of variable must be positive at any feasible solution.)

Solution For any feasible solution $\boldsymbol{x} = (x_1, ..., x_n)^\top$, denote

$$A = \{j = 1, .., n : x_j = 0\}.$$

We have $|A| \leq n-1$ since the condition $M > 0$ implies at least one of x_j is non-zero. Then, the hypersurface of active constraints is

$$\left\{\sum_{j=1}^{n} x_j = M; x_j = 0, j \in A\right\}.$$

To check the regularity, we compute the derivative of constraint functions of active constraints, and have $\nabla(x_j) = \mathbf{e}_j$ and $\nabla(\sum x_j - M) = \mathbf{1}$, where $\mathbf{e}_j$ is the vector with value 1 in its j-th entry and 0 otherwise and $\mathbf{1}$ is an all-one vector. Thus, all feasible solution are regular points since it is easy to check $\{\mathbf{1}, \mathbf{e}_j, j \in A\}$ is a set of linearly-independent vectors (see lecture 7).

(b) (5') Why is this theorem true? What does λ^* represent?

Solution

The Lagrangian of the problem is given by

$$L(x, \lambda, \mu) = \sum f_i(x_i) + \lambda(M - \sum x_i) - \sum \mu_i x_i$$

and the first order conditions will be

$$\frac{\partial L}{\partial x_i} = f_i'(x_i) - \lambda - \mu_i = 0$$

$$\lambda(M - \sum x_i) = 0$$

$$\mu_i x_i = 0$$

$$\mu_i \geq 0$$

From the first and last equation we get

$$f_i'(x_i) = \lambda + \mu_i \geq \lambda.$$

λ^* represents the rate of optimal objective change over the change of right-hand-side resource M. Note that if $x_i > 0$ then $\mu_i = 0$ which implies

$$f_i'(x_i) = \lambda$$

In this case, λ^* corresponds to the marginal gain from increasing each one of the f_i when $x_i^* > 0$, that is, the optimal allocation of x_i is at the point such that every function f_i has the same marginal gain λ^*.

(c) (5') Show that

$$\lambda^* = \frac{1}{M} \sum_j (x_j^* f_j'(x_j^*)).$$

Solution

Multiply the first FOC by x_i to get

$$f_i'(x_i)x_i - \lambda x_i - \mu_i x_i = 0$$

Replacing $\mu_i x_i = 0$ and summing over i

$$\sum f_i'(x_i)x_i = \lambda \sum x_i$$

And using the complementary condition $\lambda(M - \sum x_i) = 0$ we get

$$\sum f_i'(x_i)x_i = \lambda M$$

which gives the result.

(d) (5') What is λ^* if

$$f_j(x_j) = -w_j \log(x_j), \ \forall j,$$

where w_j is a given positive constant (the minimizer of the problem is called the weighted analytic center)?

Solution

In this case $f_i'(x_i)x_i = -w_i$ so

$$\lambda^* = -\frac{1}{M} \sum_i w_i$$

Question 4 [Small SNL] (20')

In this problem, we will solve the SDP relaxation of a small SNL problem by hand. Consider a sensor network localization problem with three anchors:

$$\mathbf{a}_1 = (1; 0),$$
$$\mathbf{a}_2 = (0; 1),$$
$$\mathbf{a}_3 = (-1; 0).$$

The sensor's true location is

$$\mathbf{x} = (1; 1.5).$$

However, the sensor's true location is unknown to the problem solver. In Figure 5, the three red points denote anchors and the blue point denotes the true location of the sensor.

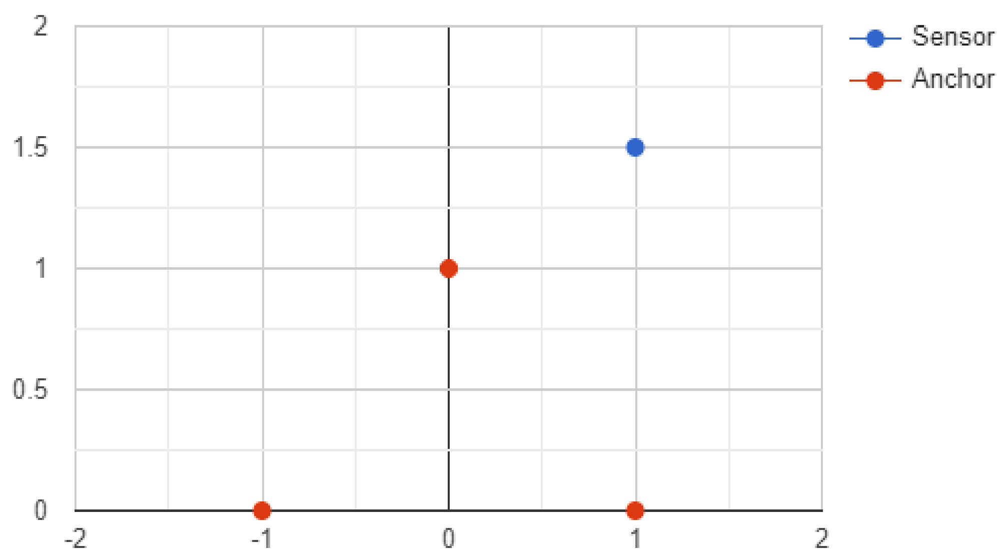

Figure 5: Location of anchors and the sensor

(a) (5') Write out the SDP relaxation problem for localizing the single sensor with null objective.

Solution We refer to Lecture 5. First, compute the distance between the sensor, and anchors and we have

$$\begin{aligned} d_1^2 &= \|\mathbf{a}_1 - \mathbf{x}\|^2 = 2.25, \\ d_2^2 &= \|\mathbf{a}_2 - \mathbf{x}\|^2 = 1.25, \\ d_3^2 &= \|\mathbf{a}_3 - \mathbf{x}\|^2 = 6.25. \end{aligned}$$

Then, the SDP relaxation problem is

$$\begin{aligned} \max_{\mathbf{Z}} \ & \mathbf{0} \cdot \mathbf{Z} \\ \text{s.t.} \ & (1;0;0)(1;0;0)^T \cdot \mathbf{Z} = 1 \\ & (0;1;0)(0;1;0)^T \cdot \mathbf{Z} = 1 \\ & (1;1;0)(1;1;0)^T \cdot \mathbf{Z} = 2 \\ & (\mathbf{a}_k; -1)(\mathbf{a}_k; -1)^T \cdot \mathbf{Z} = d_k^2, \ \text{for } k = 1,2,3 \\ & \mathbf{Z} \succeq \mathbf{0}. \end{aligned}$$

(b) (5') Write out the dual of the SDP relaxation problem.

Solution See page 13 in Lecture 5. For the primal problem,

$$
\begin{aligned}
& \max_{\mathbf{Z}} \ \mathbf{0}\cdot\mathbf{Z} \\
\text{s.t. } & (1;0;0)(1;0;0)^T\cdot\mathbf{Z} = 1 \ (w_1) \\
& (0;1;0)(0;1;0)^T\cdot\mathbf{Z} = 1 \ (w_2) \\
& (1;1;0)(1;1;0)^T\cdot\mathbf{Z} = 2 \ (w_3) \\
& (\mathbf{a}_k;-1)(\mathbf{a}_k;-1)^T\cdot\mathbf{Z} = d_k^2 \ (\lambda_k), \ \text{for } k=1,2,3 \\
& \mathbf{Z}\succeq\mathbf{0}.
\end{aligned}
$$

The dual problem is

$$
\begin{aligned}
& \min \ w_1 + w_2 + 2w_3 + \sum_{k=1}^{3}\lambda_k d_k^2 \\
\text{s.t. } & \begin{pmatrix} \begin{pmatrix} w_1+w_3 & w_3 \\ w_3 & w_2+w_3 \end{pmatrix} + \sum_{k=1}^{3}\lambda_k\mathbf{a}_k\mathbf{a}_k^T & -\sum_{k=1}^{3}\lambda_k\mathbf{a}_k \\ -\sum_{k=1}^{3}\lambda_k\mathbf{a}_k^T & \sum_{k=1}^{3}\lambda_k \end{pmatrix} \succeq \mathbf{0}.
\end{aligned}
$$

(c) (5') Write out the SDP solution explicitly constructed from the true position (slide 12 of lecture note 5) and verify that it is an optimal solution to the SDP relaxation problem.

Solution In this case,

$$
\mathbf{Z} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1.5 \\ 1 & 1.5 & 3.25 \end{pmatrix}
$$

and it is easy to verify that $\boldsymbol{Z}$ is a feasible solution of the primal SDP. Since the objective is always 0, it is also an optimal solution.

(d) (5') Verify that the dual has a rank-one optimal solution so that the SDP solution in (c) is the only optimal solution.

Solution We can find a rank-1 dual solution such that

$$\begin{pmatrix} \begin{pmatrix} w_1 + w_3 & w_3 \\ w_3 & w_2 + w_3 \end{pmatrix} + \sum_{k=1}^{3} \lambda_k \mathbf{a}_k \mathbf{a}_k^T & -\sum_{k=1}^{3} \lambda_k \mathbf{a}_k \\ -\sum_{k=1}^{3} \lambda_k \mathbf{a}_k^T & \sum_{k=1}^{3} \lambda_k \end{pmatrix} = (-\mathbf{x}; 1)(-\mathbf{x}; 1)^T.$$

Thus, any solution to the SDP relaxation is at most rank 2. Moreover, the primal problem implies that all feasible solution is at least rank 2. Thus, all primal solutions have the same rank, and the primal solution is unique based on theorem 3 in lecture 5 (To show the uniqueness, we need to verify the linear independence.)

Question 5 (10'+5'(bonus))

Consider the primal feasible region in standard form $\mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \geqslant \mathbf{0}$, where $\mathbf{A}$ is an $m \times n$ matrix, $\mathbf{b}$ is a constant nonzero m-vector, and $\mathbf{x}$ is a variable n-vector.

(a) (5') A variable x_i is said to be a *null variable* if $x_i = 0$ in every feasible solution. Prove that, if the feasible region is non-empty, x_i is a null variable if and only if there is a nonzero vector $\mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{y}^T\mathbf{A} \geq \mathbf{0},\ \mathbf{y}^T\mathbf{b} = 0$. and the ith component of $\mathbf{y}^T\mathbf{A}$ is strictly positive.

Solution:

If the feasible region is non-empty, the following LP is feasible and bounded. Then, we have that the primal optimal objective value is no larger than 0, which implies that the dual optimal objective value is also no larger than 0.

$$\begin{aligned} \text{minimize} \quad & -x_i \\ \text{subject to} \quad & \boldsymbol{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

Moreover, by Duality Theorem of Linear Programming, x_i is a null variable iff the primal optimal objective value is 0. Thus, x_i is a null variable iff there is one dual feasible solution such that the objective value is 0. This is equivalent to the condition in this question.

(b) (5') [*Strict complementarity*] Let the feasible region be nonempty. Then there is a feasible $\mathbf{x}$ and vector $\mathbf{y} \in \mathbb{R}^m$ such that

$$\mathbf{y}^T\mathbf{A} \geq \mathbf{0},\ \mathbf{y}^T\mathbf{b} = 0,\ \mathbf{A}^T\mathbf{y} + \mathbf{x} > \mathbf{0}.$$

Solution: If the feasible region is nonempty, by Part (a) we know for each $i = 1, .., n$, there exist $\boldsymbol{x}^i \in \mathbb{R}^n, \boldsymbol{y}^i \in \mathbb{R}^m$ such that

$$\boldsymbol{A}\boldsymbol{x}^i = \boldsymbol{b}, \boldsymbol{x}^i \geq \boldsymbol{0}, \boldsymbol{A}^T\boldsymbol{y}^i \geq \boldsymbol{0}, \boldsymbol{b}^T\boldsymbol{y}^i = 0,$$

and either $\boldsymbol{e}_i^T\boldsymbol{x}^i > 0$ or $\boldsymbol{e}_i^T\boldsymbol{A}^T\boldsymbol{y}^i > 0$, i.e., $\boldsymbol{e}_i^T(\boldsymbol{x}^i + \boldsymbol{A}^T\boldsymbol{y}^i) > 0$.
Let $\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}^i$ and $\bar{\boldsymbol{y}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{y}^i$, we have

$$\boldsymbol{A}\bar{\boldsymbol{x}} = \boldsymbol{b}, \bar{\boldsymbol{x}} \geq \boldsymbol{0}, \boldsymbol{A}^T\bar{\boldsymbol{y}} \geq \boldsymbol{0}, \boldsymbol{b}^T\bar{\boldsymbol{y}} = 0,$$

and $\boldsymbol{e}_i^T(\bar{\boldsymbol{x}} + \boldsymbol{A}^T\bar{\boldsymbol{y}}) > 0$ for each $i = 1, ..., n$, i.e., $\bar{\boldsymbol{x}} + \boldsymbol{A}^T\bar{\boldsymbol{y}} > \boldsymbol{0}$.

(c) (5') (Bonus) A variable x_i is a *nonextremal variable* if $x_i > 0$ in every feasible solution. Prove that, if the feasible region is non-empty, x_i is a nonextremal variable if and only if there is $\mathbf{y} \in \mathbb{R}^m$ and $\mathbf{d} \in \mathbb{R}^n$ such that $\mathbf{y}^T\mathbf{A} = \mathbf{d}^T$, where $d_i = -1,\ d_j \geqslant 0$ for $j \neq i$; and such that $\mathbf{y}^T\mathbf{b} < 0$.

Solution:

If the feasible region is non-empty, the following LP is feasible and bounded.

$$\begin{array}{ll} \text{minimize} & x_i \\ \text{subject to} & \boldsymbol{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{array}$$

Moreover, by Duality Theorem of Linear Programming, if the feasible region is non-empty, the primal problem has at least one optimal solution. It implies that x_i is a nonextremal variable iff the optimal objective value of the above primal problem is strictly positive. Thus, x_i is a nonextremal variable iff there is one dual feasible solution such that the objective value is strictly positive. That is, there exists a dual feasible solution such that

$$\boldsymbol{b}^T\boldsymbol{y} < 0,\ \boldsymbol{e}_i + \boldsymbol{A}^T\boldsymbol{y} \geq \boldsymbol{0} \tag{1}$$

which is equivalent to existence of $\boldsymbol{y}$ and $\boldsymbol{d}$ such that

$$\boldsymbol{y}^T\boldsymbol{b} < 0, \boldsymbol{y}^T\boldsymbol{A} = \boldsymbol{d}^T, d_i = -1, d_j \geq 0, \forall i \neq j,$$

Question 6 (20')

We have studied the Fisher market equilibrium problem where there are m goods in the market and each good j has a fixed amount $\bar{s}_j(>0)$ available. There are n buyers in the market where each buyer, say buyer $i \in \{1, ..., n\}$, is equipped with a fixed budget $w_i(>0)$ and independently solves a linear utility maximization problem

$$\max_{\mathbf{x}_i=(x_{i1},...,x_{im})} \quad u_i(\mathbf{x}_i) = \sum_j u_{ij} x_{ij} \quad \text{s.t. } \mathbf{p}^T \mathbf{x}_i \leq w_i, \ \mathbf{x}_i \geq \mathbf{0}.$$

Here $u_{ij} \geq 0$ is the coefficient of buyer i on good j, and decision variable x_{ij} is the amount of good j purchased by buyer i, and $\mathbf{p} \in \mathbb{R}^m$ is a given market price vector. The equilibrium prices are the prices to clear the market.

Another important utility function for each buyer i is called the Leontief utility function

$$u_i(\mathbf{x}_i) = \min_j \left\{ \frac{x_{ij}}{u_{ij}} \right\},$$

that is, the utility function value is the smallest x_{ij}/u_{ij}, $j = 1, ..., m$. For simplicity, assume $u_{ij} > 0$ for all i and j.

One fact of the Leontief utility function is that a good may not be able to be cleared (or all purchased) in the market so that its price should be zero from the economic theory. Thus, we look the equilibrium price $\mathbf{p} \in \mathbb{R}^m$ and allocation $\mathbf{x}_i \in \mathbb{R}^m$, $i = 1, ..., n$, such that the following conditions are met:

$$\sum_i x_{ij} \leq \bar{s}_j, \quad \text{and} \quad p_j \left(\bar{s}_j - \sum_{i=1}^n x_{ij} \right) = 0, \ \forall j,$$

and $\mathbf{x}_i \in \mathbb{R}^m$, $i = 1, ..., n$, is an optimal solution of

$$\max_{\mathbf{x}_i} \quad \min_j \left\{ \frac{x_{ij}}{u_{ij}} \right\} \quad \text{s.t. } \mathbf{p}^T \mathbf{x}_i \leq w_i, \ \mathbf{x}_i \geq \mathbf{0}. \tag{2}$$

(a) (5') What is the interpretation of the Leontief utility function?

Solution: (Buying Proportionally) Example: suppose x_1 is the number of left shoes and x_2 the number of right shoes; but a consumer can only use pairs of shoes. Hence, his utility is $\min\{x_1, x_2\}$. Or the cocktail example mentioned in class.

(b) (5') Write down the optimality conditions of buyer ith's maximization problem (2). (Hint: you may simplify the problem by define a scalar utility objective value z_i, then $x_{ij} = u_{ij} z_i$ so that the problem become a one-variable problem.)

Solution 1 : We can reformulate (2) as the following equivalent LP:

$$\max_{\boldsymbol{x}_i, z_i} \quad z_i \quad \text{s.t. } z_i u_{ij} \leq x_{ij}, \mathbf{p}^T \boldsymbol{x}_i \leq w_i, \ \boldsymbol{x}_i \geq \mathbf{0}. \tag{3}$$

and the optimality conditions are simply the KKT conditions

$$\begin{cases} z_i u_{ij} \leq x_{ij} \\ \mathbf{p}^T \boldsymbol{x}_i \leq w_i \\ \lambda_i p = \boldsymbol{\mu}_i + \boldsymbol{\nu}_i \\ \boldsymbol{\mu}_i^T \boldsymbol{u}_i = 1 \\ \lambda_i (\mathbf{p}^T \boldsymbol{x}_i - w_i) = 0 \\ \mu_{ij}(u_{ij} z_i - x_{ij}) = 0 \\ \nu_{ij} x_{ij} = 0 \\ \boldsymbol{x}_i, \lambda_i, \boldsymbol{\mu}_i, \boldsymbol{\nu}_i \geq \mathbf{0} \end{cases} \tag{4}$$

where λ_i is the multiplier of $\mathbf{p}^T \boldsymbol{x}_i - w_i \leq 0$, μ_{ij} is the multiplier of $z_i u_{ij} - x_{ij} \leq 0$ and ν_{ij} is the multiplier of $-x_{ij} \leq 0$. Also $\boldsymbol{u}_i$ is the length m vector with j-th component u_{ij}.

Solution 2 (5 pts): Alternatively, if we write out the dual problem

$$\begin{array}{rl} \min\limits_{\lambda_i, \boldsymbol{\mu}_i, \boldsymbol{\nu}_i} & \lambda_i w_i \\ \text{subject to} & \boldsymbol{\mu}_i^T \boldsymbol{u}_i = 1 \\ & \lambda_i \mathbf{p} = \boldsymbol{\mu}_i + \boldsymbol{\nu}_i \\ & \lambda_i, \boldsymbol{\mu}_i, \boldsymbol{\nu}_i \geq 0 \end{array} \tag{5}$$

then the equivalent optimality conditions are listed as follows (i.e. primal/dual feasibility and zero duality gap):

$$\begin{cases} \lambda_i w_i = z_i \\ \mathbf{p}^T \boldsymbol{x}_i \leq w_i \\ z_i u_{ij} \leq x_{ij} \\ \boldsymbol{\mu}_i^T \boldsymbol{u}_i = 1 \\ \lambda_i \mathbf{p} = \boldsymbol{\mu}_i + \boldsymbol{\nu}_i \\ \boldsymbol{x}_i, \lambda_i, \boldsymbol{\mu}_i, \boldsymbol{\nu}_i \geq \mathbf{0} \end{cases} \tag{6}$$

Solution 3 (5 pts): Again we can write out the dual problem and the same optimality conditions as in Solution 2, but as we know that for the primal problem to be optimal, we must have $z_i = \min_j \left\{ \frac{x_{ij}}{u_{ij}} \right\}$, hence substituting it into the above equations we have

$$\begin{cases} \lambda_i w_i = \min_j \left\{ \frac{x_{ij}}{u_{ij}} \right\} \\ \mathbf{p}^T \boldsymbol{x}_i \leq w_i \\ \lambda_i \mathbf{p} \geq \boldsymbol{\mu}_i \\ \boldsymbol{\mu}_i^T \boldsymbol{u}_i = 1 \\ \boldsymbol{x}_i, \lambda_i, \boldsymbol{\mu}_i \geq \mathbf{0} \end{cases} \tag{7}$$

All these above are equivalent, but the latter two are more useful for use in (c) and (d).

Solution 4 (5 pts) (Simplest): The simplest way is to let $x_{ij} = z_i u_{ij}$ for all j, since there is no sense to buy good j more than necessary ($x_{ij} > z_i u_{ij}$), because decreasing x_{ij} does not alter the optimal value for each individual i. Then we have individual optimality conditions as

$$x_{ij}^* = z_i^* u_{ij} \ \forall i \quad \text{where } z_i^* = \frac{w_i}{\sum_i u_{ij} p_j}; \quad \forall i.$$

One can verify that $\lambda_i = z_i / w_i$, $\boldsymbol{\mu}_i = \lambda_i \boldsymbol{p}$ are feasible.

(c) (5') Derive the equilibrium (price $\mathbf{p}$ and allocation $\mathbf{x}_i,\ i = 1, ..., n$) conditions for the Leontoef market.

Solution 1 (rigorous version): This question acts as a stepping stone for (d). The key is to prove that $\mathbf{p}^T\boldsymbol{x}_i = w_i$. Using (6), we have $\forall i = 1, \ldots, m, j = 1, \ldots, n$ that

$$\begin{aligned} & x_{ij} \geq z_i u_{ij} = u_{ij}\lambda_i w_i \geq u_{ij}\lambda_i \mathbf{p}^T\boldsymbol{x}_i \Longrightarrow \mu_{ij}x_{ij} \geq \mu_{ij}u_{ij}\lambda_i\mathbf{p}^T\boldsymbol{x}_i \\ & \Longrightarrow \boldsymbol{\mu}_i^T\boldsymbol{x}_i = \sum_{i=1}^m \mu_{ij}x_{ij} \geq \sum_{i=1}^m \mu_{ij}u_{ij}\lambda_i\mathbf{p}^T\boldsymbol{x}_i = \boldsymbol{\mu}_i^T\boldsymbol{u}_i\lambda_i\mathbf{p}^T\boldsymbol{x}_i = \lambda_i\mathbf{p}^T\boldsymbol{x}_i \end{aligned} \tag{8}$$

But we also have $\boldsymbol{\mu}_i \leq \lambda_i\mathbf{p} \Longrightarrow \lambda_i\mathbf{p}^T\boldsymbol{x}_i \geq \boldsymbol{\mu}_i^T\boldsymbol{x}_i$. Hence all the above inequalities become equalities, and hence in particular we get $\mathbf{p}^T\boldsymbol{x}_i = w_i$ and $z_i = \lambda_i w_i = \frac{x_{ij}}{u_{ij}}$ for all i and j.

Hence we have $w_i = \sum_{i=1}^m p_i x_{ij} = \sum_{i=1}^m p_i u_{ij} z_i = \mathbf{p}^T\boldsymbol{u}_i z_i$, and hence $z_i = \frac{w_i}{\mathbf{p}^T\boldsymbol{u}_i}$ and $x_{ij} = z_i u_{ij} = \frac{w_i u_{ij}}{\mathbf{p}^T\boldsymbol{u}_i}$. This solves x_{ij} and z_i out with fixed $\mathbf{p}$. And hence we can now simply write down the equilibrium conditions as follows:

$$\begin{cases} z_i = \frac{w_i}{\mathbf{p}^T\boldsymbol{u}_i} \\ x_{ij} = \frac{w_i u_{ij}}{\mathbf{p}^T\boldsymbol{u}_i} \\ \sum_j x_{ij} \leq \bar{s}_i \\ p_i\left(\bar{s}_i - \sum_{j=1}^n x_{ij}\right) = 0 \end{cases} \tag{9}$$

or equivalently

$$\begin{cases} z_i = \frac{w_i}{\mathbf{p}^T\boldsymbol{u}_i} \\ \sum_j u_{ij}z_i \leq \bar{s}_i \\ p_i\left(\bar{s}_i - \sum_{j=1}^n u_{ij}z_i\right) = 0 \end{cases} \tag{10}$$

Solution 2 (Simplest): Substitute $x_{ij} = z_i u_{ij}$ for all i and j, and then derive the equilibrium conditions using just variables z_i and p_i:

$$z_i = \frac{w_i}{\sum_i u_{ij}p_i}\ \forall j; \quad \sum_j u_{ij}z_i \leq \bar{s}_i,\ p_i\left(\bar{s}_i - \sum_{j=1}^n u_{ij}z_i\right) = 0,\ p_i \geq 0,\ \forall i.$$

Or in matrix form

$$\mathbf{z} = \mathbf{w}./U^T\mathbf{p};\ U\mathbf{z} \le \bar{\mathbf{s}},\ P(\bar{\mathbf{s}} - U\mathbf{z}) = 0,\ \mathbf{p} \ge \mathbf{0}.$$

where the ij the entry of coefficient matrix U is u_{ij} and $P = \mathrm{Diag}(\mathbf{p})$.

(d) (5') Derive a social optimization problem to represent these conditions.

Solution: The social optimization problem can be written as follows:

$$\begin{aligned}
\max_{z_i, \boldsymbol{x}_i} \quad & \sum_{j=1}^{n} w_i \log z_i \\
\text{s.t.} \quad & \sum_{j=1}^{n} x_{ij} \le \bar{s}_i \\
& z_i u_{ij} \le x_{ij},\ \ \boldsymbol{x}_i \ge 0 \\
& i = 1, \ldots, m,\ \ j = 1, \ldots, n
\end{aligned} \tag{11}$$

or the simplest (removing all $\boldsymbol{x}_i$)

$$\begin{aligned}
\max_{\mathbf{z}} \quad & \sum_{j=1}^{n} w_i \log z_i \\
\text{s.t.} \quad & \sum_{j=1}^{n} u_{ij} z_i \le \bar{s}_i \\
& i = 1, \ldots, m,\ \ j = 1, \ldots, n
\end{aligned} \tag{12}$$

This is a convex optimization problem, and hence the first-order KKT conditions are sufficient and necessary for optimality (to be rigorous, the constraint qualification is satisfied since the feasible region has an interior, but we won't require students to write down such details like checking existence of strict interior points).

Writing down the KKT conditions, we see that it's the same as what we have deduced in (c) with the simplest representation (where $\mathbf{p}$ is the multiplier of the constraints in social optimization), and hence we can solve the above the social optimization problem to get the equilibrium price and the optimal allocations altogether. □

Question 7 (Bonus 5')

Consider the unconstrained quadratic minimization problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \ f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x},$$

where $\mathbf{Q} \in \mathbb{R}^{n\times n}$ is a PSD matrix and a minimizer is $\mathbf{x}^* = \mathbf{0}$. Assume $\mathbf{Q}$ has K distinct positive eigenvalues and denote them by $\lambda_1, ..., \lambda_K$.

Let $\mathbf{x}_0$ be any initial solution in $\mathbb{R}^n$, and

$$\mathbf{x}_k = \mathbf{x}_{k-1} - \frac{1}{\lambda_k}\nabla f(\mathbf{x}_{k-1}),$$

for $k = 1, ..., K$. Show that $\mathbf{x}_K$ is one optimal solution, that is, the process stops at most K steps. (Hint: you may write $\mathbf{Q} = \sum_{k=1}^{K} \lambda_k \mathbf{v}_k \mathbf{v}_k^T$, where $\mathbf{v}_k$ is the normalized eigenvector corresponding to eigenvalue λ_k.)

Solution

Note that $\boldsymbol{x}_k = \boldsymbol{x}_{k-1} - \frac{1}{\lambda_k}\boldsymbol{Q}\boldsymbol{x}_{k-1} = (\boldsymbol{I} - \frac{\boldsymbol{Q}}{\lambda_k})\boldsymbol{x}_{k-1}$ implies that

$$\boldsymbol{x}_K = \prod_{k=1}^{K}(\boldsymbol{I} - \frac{\boldsymbol{Q}}{\lambda_k})\boldsymbol{x}_0.$$

Given the optimality condition $\boldsymbol{Q}\boldsymbol{x} = \mathbf{0}$, it suffices to prove $\boldsymbol{Q}\prod_{k=1}^{K}(\lambda_k \boldsymbol{I} - \boldsymbol{Q}) = \mathbf{0}$, which is true since $f(t) = t\prod_{k=1}^{K}(\lambda_k - t)$ can be divided by the minimal polynomial of $\boldsymbol{Q}$.

This statement also holds for general unconstrained quadratic minimization problem, i.e., $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} - \mathbf{b}^T\mathbf{x}$. Moreover, $\mathbf{Q}$ can have same eigenvalues.

In this case, if $\mathbf{b}$ is not in the column space of $\boldsymbol{Q}$, the problem is unbounded. Otherwise, we can show that $\mathbf{x}_K$ is one optimal solution.

Since $\mathbf{Q}$ is a PSD matrix, there exists an orthogonal matrix $\mathbf{A}$ such that

$$\mathbf{A}\mathbf{Q}\mathbf{A}^T = \mathbf{\Lambda},\ \mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I},$$

where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, ..., \lambda_K, 0, ..., 0)$. Since $\mathbf{Q}\mathbf{x} = \mathbf{b}$ is feasible, we have $\mathbf{b}$ is in the linear span of first K columns of A^T and perpendicular to other columns.

Let $b' = \mathbf{A}\mathbf{b}$, which is the coordinates of $\mathbf{b}$ in the basis constructed by the column space of $\mathbf{A}^T$. Then, for any given $\mathbf{x}_0$, we have

$$\begin{aligned}\mathbf{A}\mathbf{x}_1 &= \mathbf{A}\mathbf{x}_0 - \frac{1}{\lambda_1}\mathbf{A}\mathbf{Q}\mathbf{A}^T\mathbf{A}\mathbf{x}_0 + \frac{1}{\lambda_1}\mathbf{A}\mathbf{b} \\ &= \mathbf{A}\mathbf{x}_0 - \frac{1}{\lambda_1}\mathbf{\Lambda}\mathbf{A}\mathbf{x}_0 + \frac{1}{\lambda_1}\mathbf{A}\mathbf{b}.\end{aligned}$$

It implies that $(\mathbf{A}\mathbf{x}_1)_1 = \frac{1}{\lambda_1}b'_1$. Moreover, by induction, we can show that for any $k \geq 1$, if $(\mathbf{A}\mathbf{x}_k)_1 = \frac{1}{\lambda_1}b'_1$, it also holds for $k+1$. Thus, we have $\lambda_1(\mathbf{A}\mathbf{x}_k) = b'_1$ for all k.

Furthermore, similarly, we can show that at step k, the gradient descent actually find one solution such that $\lambda_k(\mathbf{A}\mathbf{x}_k)_k = b'_k$, i.e., identifying the coordinates of $\mathbf{b}$ in the basis constructed by the column space of A^T. After K steps, we have

$$\mathbf{A}\mathbf{x}_K = \mathbf{\Lambda}^{-1}\mathbf{A}\mathbf{b} \Rightarrow \mathbf{Q}\mathbf{x}_K = \mathbf{b}.$$