Optimization Methods for Data Science, Machine Learning and AI
Prof. Yinyu Ye

Homework Assignment 2 Sample Solution

Individual Homework (110'):

1. (15') Consider problem 5 of Homework Assignment 1 where the second-order cone is replaced by the p-th order cone for $p \geq 1$:

$$\begin{array}{rl} \min_{\mathbf{x}} & 2x_1 + x_2 + x_3 \\ \text{s.t.} & x_1 + x_2 + x_3 = 1, \\ & x_1 - \|(x_2,\ x_3)\|_p \geq 0. \end{array}$$

   (a) (5') Write out the conic dual problem.

   (b) (5') Compute the dual optimal solution $(y^*, \mathbf{s}^*)$.

   (c) (5') Using the zero duality condition to compute the primal optimal solution $\mathbf{x}^*$.

Solution:

(a) Following lecture note 3, slide 19, the dual is

$$\max\ y \quad \text{s.t.} \quad ye + s = (2, 1, 1)^T,\ s_1 - \|(s_2,\ s_3)\|_q \geq 0$$

or

$$\max\ y \quad \text{s.t.} \quad (2 - y) - (2|1 - y|^q)^{1/q} \geq 0$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

(b) If $y \geq 1$, the constraint can be written as $(2 - y) - 2^{1/q}(y - 1) \geq 0$ so that the maximal value is

$$y^* = \frac{2 + 2^{1/q}}{1 + 2^{1/q}}.$$

which is indeed ≥ 1. Hence there is no need to consider the other case when $y < 1$. And $s^* = (2 - y^*; 1 - y^*; 1 - y^*)^T$. For $p = 1$, $y^* = 3/2$; $p = 2$, $y^* = \sqrt{2}$; and for $p = \infty$, $y^* = 4/3$.

(c) From the zero duality condition, we have $2x_1^* + x_2^* + x_3^* = y^*$, and together with the constraints $x_1^* + x_2^* + x_3^* = 1$, we have

$$x_1^* = y^* - 1 = \frac{1}{1 + 2^{1/q}}, \quad x_2^* + x_3^* = \frac{2^{1/q}}{1 + 2^{1/q}}.$$

When $x_2^* = x_3^* = \frac{2^{1/q}}{2(1+2^{1/q})} > 0$,

$$\|(x_2^*; x_3^*)\|_p^p = 2\left(\frac{2^{1/q}}{2(1+2^{1/q})}\right)^p = \frac{1}{(1+2^{1/q})^p}2^{1-p+p/q} = \frac{1}{(1+2^{1/q})^p} \le (x_1^*)^p$$

so that it is feasible and, consequently, optimal.

This optimal solution is also unique, as we have

$$\frac{2^{1/q}}{1+2^{1/q}} = x_2^* + x_3^* \le \|(x_2^*; x_3^*)\|_p \|(1; 1)\|_q = 2^{1/q}\|(x_2^*; x_3^*)\|_p$$

by Holder's inequality, which implies that

$$\|(x_2^*; x_3^*)\|_p \ge \frac{1}{1+2^{1/q}} = x_1^*$$

and the equality is obtained iff $x_2^* = x_3^* = \frac{2^{1/q}}{2(1+2^{1/q})}$.

2. (20') Consider the distributionally robust optimization (DRO) problem

$$\text{minimize}_{\mathbf{x}\in X} \quad \left[\max_{\mathbf{d}\in D} \quad \sum_{k=1}^N (\hat{p}_k + d_k) h(\mathbf{x}, \xi_k)\right] \tag{1}$$

where the distribution set D is now given by

$$D = \{\mathbf{d} : \sum_{k=1}^N d_k = 0, \ \|\mathbf{d}\|^2 \le 1/N, \ \hat{p}_k + d_k \ge 0, \ \forall k.\}$$

(a) (3') What is the interpretation of D? Answer within 2 sentences.

(b) (4') Represent D in standard conic form. (Hint: one set of the slack variables are in the second-order cone and the others are in the non-negative orthant cone.)

(c) (7') Construct the conic dual of the inner max-problem.

(d) (6') Replace the inner max-problem (1) by its dual, and simplify the DRO problem as much as possible.

Solution:

(a) D denotes a set of bounded perturbations $\mathbf{d}$ (or slack variables) which keep the resulting $p_k := \hat{p}_k + d_k$, $k = 1, \ldots, N$ a probability vector.

(b) The conic representation of D is

$$\left\{(d_0; \mathbf{d}) : d_0 = 1/\sqrt{N}, \sum_{k=1}^N d_k = 0, \hat{p}_k + d_k = p_k, p_k \ge 0, \|\mathbf{d}\| \le d_0\right\}$$

(c) Denoting $\mathbf{h} := (h(x,\xi_1);\ldots;h(x,\xi_N))$ and $\hat{\mathbf{p}} := (\hat{p}_1;\ldots;\hat{p}_N)$, and ignoring the constants $\sum_{k=1}^N \hat{p}_k h(x,\xi_k)$, the primal problem can be abbreviated as the following CLP:

$$\begin{aligned}\min_{d_0;\mathbf{d};\mathbf{y}} \quad & -\mathbf{h}^T\mathbf{d} \\ \text{s.t.} \quad & d_0 = 1/\sqrt{N}, e^T\mathbf{d} = 0, \mathbf{d} - \mathbf{y} = -\hat{\mathbf{p}} \\ & (d_0;\mathbf{d}) \in SOC^{N+1}, \mathbf{y} \geq 0\end{aligned}$$

Suppose that $\lambda_0, \lambda_1, \lambda_2$ are the multipliers for the corresponding equality constraints, then the dual problem is

$$\begin{aligned}\min_{\lambda_0,\lambda_1,\lambda_2} \quad & \lambda_0/\sqrt{N} - \lambda_2^T\hat{\mathbf{p}} \\ \text{s.t.} \quad & \lambda_0(1;0;0) + \lambda_1(0;e;0) + \lambda_2(0;I;-I) - (s_0;\mathbf{s};\mathbf{z}) = (0;\mathbf{h};0) \\ & (s_0;\mathbf{s}) \in SOC^{N+1}, \mathbf{z} \geq 0\end{aligned}$$

or equivalently,

$$\begin{aligned}\min_{\lambda_0,\lambda_1,\lambda_2} \quad & \lambda_0/\sqrt{N} - \lambda_2^T\hat{\mathbf{p}} \\ \text{s.t.} \quad & \|\lambda_1 e + \lambda_2 - \mathbf{h}\| \leq \lambda_0 \\ & \lambda_2 \leq 0\end{aligned}$$

which can be further simplified to

$$\begin{aligned}\min_{\lambda_1,\lambda_2} \quad & \|\lambda_1 e + \lambda_2 - \mathbf{h}\|/\sqrt{N} - \lambda_2^T\hat{\mathbf{p}} \\ \text{s.t.} \quad & \lambda_2 \leq 0\end{aligned}$$

(d) Replacing the inner-max problem with its dual in (c), we can reformulate the DRO problem as follows:

$$\begin{aligned}\min_{x\in X,\lambda_1,\lambda_2} \quad & \hat{\mathbf{p}}^T\mathbf{h} + \|\lambda_1 e + \lambda_2 - \mathbf{h}\|/\sqrt{N} - \lambda_2^T\hat{\mathbf{p}} \\ \text{s.t.} \quad & \lambda_2 \leq 0\end{aligned}$$

where $\hat{\mathbf{p}}$ and $\mathbf{h}$ are defined as in (c). When $x \in X$ and $\lambda_2 \leq 0$ are fixed, λ_1 can be partially solved out as

$$\lambda_1 = \frac{1}{N}\sum_{k=1}^N (h(x,\xi_k) - \lambda_2^k) = e^T(\mathbf{h} - \lambda_2)/N$$

and hence we finally arrive at

$$\begin{aligned}\min_{x\in X,\lambda_2} \quad & \hat{\mathbf{p}}^T\mathbf{h} + \|H_n(\mathbf{h} - \lambda_2)\|/\sqrt{N} - \lambda_2^T\hat{\mathbf{p}} \\ \text{s.t.} \quad & \lambda_2 \leq 0\end{aligned}$$

where $H_n := I - \dfrac{ee^T}{N}$ is the centralization matrix.

3. (10') Consider the SOCP relaxation in problem 8 of Homework Assignment 1:

$$\begin{array}{ll} \min_{\mathbf{x}} & \mathbf{0}^T\mathbf{x} \\ \text{s.t.} & \|\mathbf{x}-\mathbf{a}_i\|^2 \le d_i^2,\ i=1,2,3, \end{array}$$

where $\mathbf{x} \in R^2$.

(a) (4') Write down the first-order KKT optimality conditions.

(b) (3') Interpret (with no more than 2 sentences) the three optimal multipliers when the true position of the sensor is inside the convex hull of the three anchors.

(c) (3') Could the true position $\bar{\mathbf{x}} \in R^2$ of the sensor satisfy the optimality conditions if it is outside the convex hull of the three anchors? What would be the multiplier values?

Solution: Let the Lagrangian or dual multipliers be $y_i \le 0$, $i = 1, 2, 3$.

(a) Then, writing down the (first-order) KKT conditions, the optimal solution would satisfy

$$\sum_i y_i(\mathbf{x}-\mathbf{a}_i) = 0,$$

and complementarity

$$y_i(d_i^2 - \|\mathbf{x}-\mathbf{a}_i\|^2) = 0,\ i = 1, 2, 3.$$

(b) When the true position $\bar{\mathbf{x}} \in R^2$ is inside the convex hull, then y_i represents a force pulling $\bar{\mathbf{x}}$ from a_i. The three forces balance at $\bar{\mathbf{x}}$ as the conditions indicated. In particular, when y_i's are not all zero, then we have $\bar{\mathbf{x}} = \frac{y_1\mathbf{a}_1+y_2\mathbf{a}_2+y_3\mathbf{a}_3}{y_1+y_2+y_3}$.
Moreover, if all the forces are nonzero, then we find the correct solution. This is because the complementarity conditions then indicate that each constraint is tight, that is,

$$d_i^2 - \|\mathbf{x}-\mathbf{a}_i\|^2 = 0,\ \forall i = 1, 2, 3$$

which mean that you find the x that satisfies all the original equality constraints. In this case, the relaxation is exact.

(c) It still satisfies the optimality conditions. But all multipliers must have 0 values, since otherwise we will have $\bar{\mathbf{x}} = \frac{y_1\mathbf{a}_1+y_2\mathbf{a}_2+y_3\mathbf{a}_3}{y_1+y_2+y_3}$ with $y_i \le 0$, which is a point inside the convex hull. This leads to a contradiction. In this case, the $\mathbf{x}$ you find may not have all the constraints active, i.e.

$$d_i^2 - \|\mathbf{x}-\mathbf{a}_i\|^2 = 0,\ \forall i = 1, 2, 3$$

may not all hold.

4. (10') Consider the following parametric QCQP problem for a parameter $\kappa > 0$:

$$\begin{aligned} \min \quad & (x_1 - 1)^2 + x_2^2 \\ \text{s.t.} \quad & -x_1 + \tfrac{x_2^2}{\kappa} \geq 0 \end{aligned}$$

(a) (5') Is $\mathbf{x} = \mathbf{0}$ a first-order KKT solution?

(b) (5') Is $\mathbf{x} = \mathbf{0}$ a second-order KKT necessary or sufficient solution for some value of κ?

Solution: Define $f(x) := (x_1 - 1)^2 + x_2^2$, $c(x) = -x_1 + \frac{x_2^2}{\kappa}$. Then the Lagrangian function for this problem is

$$L(x, y) = f(x) - yc(x) = (x_1 - 1)^2 + x_2^2 - y\left(-x_1 + \frac{x_2^2}{\kappa}\right), \ y \geq 0.$$

(a) Firstly, $x = 0$ is feasible with $c(x) = 0$. Moreover,

$$\nabla f(0) = (-2;\ 0), \quad \nabla c(0) = (-1;\ 0)$$

Thus $y = 2$ makes $\nabla f(0) = 2\nabla c(0)$ so that $x = 0$ is a first-order KKT solution.

(b) Since the constraint is active, the tangent space is

$$T = \{\mathbf{d} : \ \mathbf{d} \in R^2, \ (-1, \ 0)\mathbf{d} = 0\}.$$

The second-order necessary condition implies that for all $d \in T$

$$d^T \nabla_x^2 L(\bar{x}, \bar{y}) d \geq 0,$$

where

$$\nabla_x^2 L(0, 2) = \begin{pmatrix} 2 & 0 \\ 0 & 2 - \frac{4}{\kappa} \end{pmatrix}$$

Thus, when $\kappa \geq 2$, the Hessian matrix of the Lagrangian is PSD so that $x = 0$ is a second-order KKT solution. Otherwise, $x = 0$ cannot be a local minimizer.

5. (20') (Central-Path and Potential) Given standard LP problem

$$\begin{aligned} \text{minimize}_{\mathbf{x} \in R^n} \quad & \mathbf{c}^T \mathbf{x} \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{LP}$$

The **Analytic Center** of the primal feasible region $\mathcal{F}_p := \{\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ is defined as the solution of the following linear-constrained convex optimization problem:

$$\begin{aligned} \text{minimize}_{\mathbf{x} \in R^n} \quad & -\sum_{j=1}^{n} \log x_j, \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} > \mathbf{0}. \end{aligned} \tag{PB}$$

The **Central Path** $\mathbf{x}(\mu)$ of (LP) is defined as the solution of the following Barrier LP problem (where $\mu > 0$ is a parameter):

$$\begin{aligned} \text{minimize}_{\mathbf{x}\in R^n} \quad & \mathbf{c}^T\mathbf{x} - \mu \cdot \sum_{j=1}^{n} \log x_j, \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} > \mathbf{0}. \end{aligned} \tag{BLP}$$

Part I Now consider the following example:

$$\begin{aligned} \text{minimize}_{\mathbf{x}\in R^3} \quad & x_1 + x_2, \\ \text{subject to} \quad & x_1 + x_2 + x_3 = 1, \\ & (x_1, x_2, x_3) \geq \mathbf{0}. \end{aligned} \tag{2}$$

(a) (4') What is the analytic center of the primal feasible region in (2)?

(b) (4') Find the central path $\mathbf{x}(\mu) = (x_1(\mu), x_2(\mu), x_3(\mu))$ for (2).

(c) (4') Show that as μ decreases to 0, $\mathbf{x}(\mu)$ converges to the unique optimal solution of (2).

Part II Consider another example with different objective but the same feasible region:

$$\begin{aligned} \text{minimize}_{\mathbf{x}\in R^3} \quad & x_1 \\ \text{subject to} \quad & x_1 + x_2 + x_3 = 1 \\ & (x_1, x_2, x_3) \geq \mathbf{0} \end{aligned} \tag{3}$$

(d) (4') Find the central path $\mathbf{x}(\mu) = (x_1(\mu), x_2(\mu), x_3(\mu))$ for (3).

(e) (4') Which point does the central path converge to now (as $\mu \to 0+$)?

Solution:

(a) The analytic center is the vector that minimizes the potential function:

$$-\sum_{j=1}^{3} \log x_j$$

and satisfies $\sum_{j=1}^{3} x_j = 1$, $\mathbf{x} > 0$. Thus the analytic center is $(1/3, 1/3, 1/3)$.

(b) From the central path condition we derive a quadratic equation for x_1:

$$2(x_1)^2 - (3\mu + 1)x_1 + \mu = 0.$$

Taking the non-negative root gives

$$x_1 = \frac{3\mu + 1 - \sqrt{9\mu^2 + 1 - 2\mu}}{4}.$$

Other conditions give $x_2 = x_1$ and $x_3 = 1 - 2x_1$.

(c) The set of optimal solution is a singleton $(0;0;1)$. When μ decreases to zero, we know from the expression that $x_1(\mu) = x_2(\mu) \to 0$. Also, since $\sum_i x_1 = 1$ always holds, we know that $x_3 \to 1$. We know that $(0,0,1)$ is going to be the optimal solution, because $f(x) = x_1 + x_2 \geq 0$, and $(0,0,1)$ attains the value 0. The uniqueness is easily proved: to attain optimal value, x_1, x_2 has to be zero, so x_3 have to be 1, because of the equality constraint.

Thus, as μ goes to zero, $x(\mu)$ converges to the unique optimal solution.

(d)(e) Just repeat the above stuff. The only thing to be noted is that now the optimal solution to the original problem is **not unique**, so the problem description in (c) needs to be slightly changed. But everything else is the same.

6. (15') Consider the following SVM problem, where $\mu \geq 0$ is a prescribed constant:

$$\begin{array}{ll} \min & \beta + \mu\|\mathbf{x}\|^2 \\ \text{s.t.} & a_i^T\mathbf{x} + x_0 + \beta \geq 1, \ \forall i, \\ & b_j^T\mathbf{x} + x_0 - \beta \leq -1, \ \forall j, \\ & \beta \geq 0. \end{array}$$

(a) (8') Write out the Lagrangian dual problem of the SVM problem. Write it as explicit as possible (at least remove the inner minimization). (Hint: You may want to consider two separate cases: $\mu = 0$ and $\mu > 0$)

(b) (7') Suppose that we have 6 training data in R^2: $a_1 = (0;0),\ a_2 = (1;0),\ a_3 = (0;1)$ and $b_1 = (0;0),\ b_2 = (-1;0),\ b_3 = (0;-1)$. Use the optimality conditions (or any approach you want) to find optimal solutions for $\mu = 0$ and $\mu = 10^{-5}$, respectively. Are the two optimal solutions unique for the given μ? Prove your claim.

Solution:

(a) Let the multipliers for a_i constraints be $y_i^a \geq 0$ and those for b_j constraints be $y_j^b \leq 0$, and $\beta \geq 0$ be $y^\beta \geq 0$. Then, the Lagrangian function is

$$L(x, x_0, \beta, y^a, y^b, y^\beta) = \beta + \mu\|x\|^2 - \sum_i y_i^a(a_i^T x + x_0 + \beta - 1) - \sum_j y_j^b(b_j^T x + x_0 - \beta + 1) - y^\beta \beta.$$

The dual must have constraint

$$\sum_i y_i^a + \sum_j y_j^b = 0$$

and

$$1 - y^\beta - \sum_i y_i^a + \sum_j y_j^b = 0,$$

since otherwise the primal can choose x_0 or β to make the Lagrangian function unbounded from below.

1) If $\mu = 0$, then we also have

$$\sum_i y_i^a a_i + \sum_j y_j^b b_j = 0,$$

since otherwise the primal can choose x to make the Lagrangian function unbounded from below.

The dual problem is thusly

$$\begin{array}{rl} \max & \sum_i y_i^a - \sum_j y_j^b, \\ \text{s.t.} & \sum_i y_i^a + \sum_j y_j^b = 0, \\ & 1 - y^\beta - \sum_i y_i^a + \sum_j y_j^b = 0, \\ & \sum_i y_i^a a_i + \sum_j y_j^b b_j = 0, \\ & y^a \geq 0, \ y^b \leq 0, \ y^\beta \geq 0. \end{array}$$

2) For $\mu > 0$, the primal minimization of the Lagrangian function would be $\beta = 0$ and

$$2\mu x = \sum_i y_i^a a_i + \sum_j y_j^b b_j.$$

Thus,

$$\phi(y^a, y^b, y^\beta) = -\frac{1}{4\mu} \| \sum_i y_i^a a_i + \sum_j y_j^b b_j \|^2 + \sum_i y_i^a - \sum_j y_j^b,$$

and the dual problem is

$$\begin{array}{rl} \max & \phi(y^a, y^b, y^\beta) \\ \text{s.t.} & \sum_i y_i^a + \sum_j y_j^b = 0, \\ & 1 - y^\beta - \sum_i y_i^a + \sum_j y_j^b = 0, \\ & y^a \geq 0, \ y^b \leq 0, \ y^\beta \geq 0. \end{array}$$

(b) Firstly, we show that for the set of a_i, b_j given in this problem, any feasible β satisfies $\beta \geq 1$. To see this, suppose on the contrary that $\beta < 1$. Then for $a_1 = b_1$, we have

$$a_1^T \mathbf{x} + x_0 \geq 1 - \beta > 0 > -1 + \beta \geq b_1^T \mathbf{x} + x_0,$$

which is a contradiction. Hence the optimal value $\beta + \mu \|x\|^2$ of the primal objective function is at least 1. Moreover, it can always be achieved by simply setting $\beta = 1$, $\mathbf{x} = \mathbf{0}$ and $x_0 = 0$. Hence we know that the optimal value is always 1 no matter whether $\mu = 0$ or not.

1) For $\mu = 0$, any point of the form $\beta = 1$, $\mathbf{x} = (t; t)$, $x_0 = 0$ with $t \geq 0$ is optimal, as the objective value is 1 and the constraints are satisfied. So the optimal solution is not unique.

2) For $\mu > 0$, a point is optimal iff $\beta = 1$ and $\mathbf{x} = \mathbf{0}$, since otherwise we will have $\beta + \mu\|x\|^2 > \beta \geq 1$. In this case, we need $x_0 \geq 0 \geq x_0$, and hence $x_0 = 0$. Hence we obtain a unique optimal solution $\beta = 1$, $\mathbf{x} = \mathbf{0}$ and $x_0 = 0$.

7. (20') Consider a generalized Arrow–Debreu equilibrium problem in which the market has n agents and m goods. Agent i, $i = 1, ..., n$, has a bundle amount of $\mathbf{w}_i = (w_{i1}, w_{i2}, \ldots, w_{im}) \in R^m_+$ goods initially and has a linear utility function whose coefficients are $\mathbf{u}_i = (u_{i1}, u_{i2}, \ldots, u_{im}) > 0 \in R^m$. The goal is to price each good so that the market clears. Note that, given the price vector $\mathbf{p} = (p_1, p_2, \ldots, p_m) > 0$, agent i's utility maximization problem is:

$$\begin{array}{rl} \text{maximize} & \mathbf{u}_i^T \mathbf{x}_i \\ \text{subject to} & \mathbf{p}^T \mathbf{x}_i \leq \mathbf{p}^T \mathbf{w}_i \\ & \mathbf{x}_i \geq 0 \end{array}$$

(a) (5') For a given $\mathbf{p} \in R^m$, write down the optimality conditions for agent i's utility maximization problem. Without loss of generality, you may fix $p_m = 1$ since the budget constraints are homogeneous in p.

(b) (5') Suppose that $\mathbf{p} \in R^m$ and $\mathbf{x}_i \in R^m$ satisfy the constraints:

$$\begin{aligned} \sum_{i=1}^n \mathbf{x}_i &= \sum_{i=1}^n \mathbf{w}_i, \\ \frac{\mathbf{u}_i^T \mathbf{x}_i}{\mathbf{p}^T \mathbf{w}_i} p_j &\geq u_{ij}, \quad \forall i, j, \\ \mathbf{p} &\geq \mathbf{0}, \\ \mathbf{x}_i &\geq \mathbf{0}, \quad \forall i. \end{aligned}$$

Show that $\mathbf{p}$ is then an equilibrium price vector.

(c) (5') For simplicity, assume that all u_{ij} are positive so that all p_j are positive. By introducing new variables $y_j = \log(p_j)$ for $j = 1, ..., m$, the conditions can be written as follows:

$$\begin{array}{lll} \min & 0 & \\ \text{s.t.} & \sum_{i=1}^n \mathbf{x}_i = \sum_{i=1}^n \mathbf{w}_i & \\ & \log(\mathbf{u}_i^T \mathbf{x}_i) - \log\left(\sum_{k=1}^m w_{ik} e^{y_k}\right) + y_j \geq \log(u_{ij}) & \forall i, j \\ & x_{ij} \geq 0, & \forall i, j \end{array}$$

Show that this problem is convex in x_{ij} and y_j. (Hint: Use the fact that $\log\left(\sum_{k=1}^m w_{ik} e^{y_k}\right)$ is a convex function in the y_k's.)

(d) (5') Consider the Fisher example on Lecture Note with two agents and two goods, where the utility coefficients are given by

$$\mathbf{u}_1 = (2;\ 1) \quad \text{and} \quad \mathbf{u}_2 = (3;\ 1),$$

while now there are no fixed budgets. Rather, let

$$\mathbf{w}_1 = (1;\ 0) \quad \text{and} \quad \mathbf{w}_2 = (0;\ 1)$$

that is, agent 1 brings in one unit good x and agent brings in one unit of good y. Find the Arrow–Debreu equilibrium prices, where you may assume $p_y = 1$.

Solution:

(a) Notice that here p is fixed, and hence the problem is simply an LP. Writing down the primal feasibility, dual feasibility and zero duality gap conditions, we obtain:

$$u_i \le \lambda_i p, \quad \lambda_i \ge 0, \quad \lambda_i \cdot p^T w_i = u_i^T x_i, \quad x_i \ge 0, \quad p^T x_i \le p^T w_i.$$

Alternative solution: write down the KKT conditions – the zero duality gap condition $\lambda_i \cdot p^T w_i = u_i^T x_i$ will be replaced by the zero gradient condition for the Lagrangian. Notice that these are equivalent.

(b) This proof is identical to the Lecture Note #5 for Fisher equilibrium where scalar w_i is substituted by $p^T w_i$.

In particular, we simply check that x_i are all optimal for the given p in their own utility maximization LPs, i.e. we check that the optimality conditions in (a) are all satisfied.

Firstly, define $\lambda_i := \frac{u_i^T x_i}{p^T w_i}$. Then obviously we have $\lambda_i \ge 0$ and $\lambda_i p \ge u_i$ by the second set of constraints in (b). Moreover, by definition, we have $\lambda_i p^T w_i = u^T x_i$, and $x_i \ge 0$ is satisfied automatically by the third set of constraints in (b).

It remains to check that $p^T x_i \le p^T w_i$. To see this, multiply both sides of the first set of constraints in (b) by p^T, we have

$$\sum_{i=1}^{n} p^T x_i = \sum_{i=1}^{n} p^T w_i$$

On the other hand, multiplying both sides of the second set of constraints in (b) by x_j and sum over j, we have

$$\frac{u_i^T x_i}{p^T w_i} p^T x_i \ge u_i^T x_i$$

and since $u_i > 0$ by assumption, we have $\frac{u_i^T x_i}{p^T w_i}, p_j$ both strictly large than 0 (since otherwise the second set of constraints in (b) would be violated). In particular, we have $u_i^T x_i > 0$, and hence we can divide it on both sides of the above inequality, and obtain that $p^T x_i \ge p^T w_i$. Combining this with the fact that $\sum_{i=1}^n p^T x_i = \sum_{i=1}^n p^T w_i$, we conclude that $p^T w_i = p^T x_i \ge p^T x_i$, which finishes our proof. $\square$

(c) We first observe that the function $\log(u_i^T x_i)$ is concave in x_i, and that the function $g : R^m \to R$ given by $g(y) = \log\left(\sum_{k=1}^m w_{ik} e^{y_k}\right)$ is convex in y. The former is obvious. To establish the latter, we compute:

$$\frac{\partial g}{\partial y_j} = \frac{w_{ij} e^{y_j}}{S} \qquad \text{where } S = \sum_{k=1}^m w_{ik} e^{y_k}$$

$$\frac{\partial^2 g}{\partial y_j y_k} = \frac{S w_{ij} e^{y_j} \mathbf{1}_{\{j=k\}} - w_{ij} w_{ik} e^{y_j} e^{y_k}}{S^2}$$

(Optional) We show that the Hessian matrix $\nabla^2 g(y)$ is positive semidefinite by showing that it is symmetric *diagonally dominant*, and that its diagonal entries are non–negative. The symmetry of $\nabla^2 g(y)$ is obvious. Now, for all $j = 1, \ldots, m$, we have:

$$\sum_{k:k\neq j} \left| \frac{\partial^2 g}{\partial y_j y_k} \right| = \frac{1}{S^2} w_{ij} e^{y_j} \sum_{k:k\neq j} w_{ik} e^{y_k} = \frac{1}{S^2} w_{ij} e^{y_j} \left(S - w_{ij} e^{y_j}\right) = \frac{\partial^2 g}{\partial y_j^2}$$

i.e. $\nabla^2 g(y)$ is diagonally dominant. Moreover, since $w_i \geq 0$ for all $i = 1, \ldots, n$, we have:

$$\frac{\partial^2 g}{\partial y_j^2} = \frac{1}{S^2}\left(S w_{ij} e^{y_j} - w_{ij}^2 e^{2y_j}\right) = \frac{1}{S^2} \sum_{k:k\neq j} w_{ik} e^{y_k} \geq 0$$

for all $j = 1, \ldots, m$. It follows that $\nabla^2 g(y) \succeq 0$, which in turn implies that g is convex. Hence, we conclude that the inequalities:

$$\log\left(\sum_{k=1}^m w_{ik} e^{y_k}\right) - \log\left(u_i^T x_i\right) - y_j \leq -\log(u_{ij}) \quad \forall i, j$$

define a convex set. As the remaining constraints and the objective function are linear, we conclude that the problem is a convex minimization problem.

(d) The problem reduces to finding $p_x, x_1, y_1, x_2, y_2 \geq 0$ with $p_y = 1$, such that:

$$\begin{aligned}
x_1 + x_2 &= 1 \\
y_1 + y_2 &= 1 \\
\frac{2x_1 + y_1}{p_x} p_x &\geq 2 \\
\frac{2x_1 + y_1}{p_x} p_y &\geq 1 \\
\frac{3x_2 + y_2}{p_y} p_x &\geq 3 \\
\frac{3x_2 + y_2}{p_y} p_y &\geq 1
\end{aligned}$$

.

Then, you will find (either by taking a guess from the numerical solutions, or follow some case-by-case analysis arguments that we will elaborate on in the coming problem session on Feb. 16):

$$p_x = 2,\ p_y = 1,\ x_1 = 1/2,\ y_1 = 1,\ x_2 = 1/2,\ y_2 = 0.$$

8. First, we reformulate this problem in a standard SDP form. Since $\mathbf{A} = \{a_{ij}\}_{i,j=1}^{3}$ is not a zero matrix, we first assume $a_1 1 \neq 0$. Then, the dual problem can be reformulated as

$$\begin{aligned}
&\min \langle \mathbf{S} - \mathbf{C}, b\mathbf{e}_{11} \rangle \\
&\text{s.t. } \langle \mathbf{S} - \mathbf{C}, a_{11}\mathbf{e}_{11} - a_{ij}\mathbf{e}_{ij}/a_{11} \rangle = 0,\ \forall 1 \leq i \leq j \leq 3, \\
&\qquad \mathbf{S} \succeq 0,
\end{aligned}$$

where $\mathbf{e}_{ij}$ is a matrix with value 1 at (i, j) entry and zero otherwise. Here, the first constraint comes from $\mathbf{A}y + \mathbf{S} = \mathbf{C}$. Next, we apply Caratheodory's theorem to draw the conclusion. Notice that the condition is satisfies automatically when $i = j = 1$. We eliminate this constraint, and this new SDP problem only has 5 constraints. By Caratheodory's theorem, the rank r of one optimal solution satisfies

$$r(r + 1) \leq 10,$$

which implies $r \leq 2$.

Moreover, the location of the non-zero entry of $\mathbf{A}$ will not affect the following proof. Thus, we finish the proof.

Groupwork (40') (group of 1-4 people):

9. (5') Let $\{(\mathbf{a}_i, c_i)\}_{i=1}^m$ be a given dataset where $\mathbf{a}_i \in R^n$, $c_i \in \{\pm 1\}$. In Logistic Regression (LR), we determine $x_0 \in R$ and $\mathbf{x} \in R^n$ by maximizing

$$\left(\prod_{i,c_i=1} \frac{1}{1+\exp(-\mathbf{a}_i^T\mathbf{x} - x_0)}\right)\left(\prod_{i,c_i=-1} \frac{1}{1+\exp(\mathbf{a}_i^T\mathbf{x} + x_0)}\right).$$

which is equivalent to maximizing the log-likelihood probability

$$-\sum_{i,c_i=1} \log\left(1+\exp(-\mathbf{a}_i^T\mathbf{x} - x_0)\right) - \sum_{i,c_i=-1} \log\left(1+\exp(\mathbf{a}_i^T\mathbf{x} + x_0)\right).$$

In this problem, we consider the quadratic regularized log-logistic-loss function

$$f(\mathbf{x}, x_0) = \sum_{i,c_i=1} \log\left(1+\exp(-\mathbf{a}_i^T\mathbf{x} - x_0)\right) + \sum_{i,c_i=-1} \log\left(1+\exp(\mathbf{a}_i^T\mathbf{x} + x_0)\right) + 0.001\cdot\|\mathbf{x}\|_2^2.$$

Consider the following data set

$$\mathbf{a}_1 = (0;0),\quad \mathbf{a}_2 = (1;0),\quad \mathbf{a}_3 = (0;1),\quad \mathbf{a}_4 = (0;0),\quad \mathbf{a}_5 = (-1;0),\quad \mathbf{a}_6 = (0;-1),$$

with label

$$c_1 = c_2 = c_3 = 1,\quad c_4 = c_5 = c_6 = -1$$

use the KKT conditions to find a solution of min $f(\mathbf{x}, x_0)$. You can either solve it numerically (e.g., using MATLAB `fsolve`) or analytically (represent the solution by a solution of a simpler (1D) nonlinear equation).

Solution:

Since the problem is unconstrained, the KKT condition is nothing but setting $\nabla f(\mathbf{x}, x_0)$ to zero. Let $\mathbf{x} = (x_1; x_2)$, the KKT condition can be written coordinate-wise as

$$\begin{aligned}
0 &= \frac{-1}{1+\exp(x_0)} + \frac{-1}{1+\exp(x_0+x_1)} + \frac{-1}{1+\exp(x_0+x_2)} + \frac{1}{1+\exp(-x_0)} + \frac{1}{1+\exp(-x_0+x_1)} + \frac{1}{1+\exp(-x_0+x_2)} \\
0 &= -\frac{1}{1+\exp(x_0+x_1)} - \frac{1}{1+\exp(-x_0+x_1)} + 0.002x_1 \\
0 &= -\frac{1}{1+\exp(x_0+x_2)} - \frac{1}{1+\exp(-x_0+x_2)} + 0.002x_2
\end{aligned} \tag{4}$$

Note that if $x_0 = 0$ then the first equation of (4) automatically holds. Assuming $x_0 = 0$, the last two equations becomes

$$x_1 = \frac{1000}{1+\exp(x_1)},\quad x_2 = \frac{1000}{1+\exp(x_2)}$$

Hence it suffices to set $x_1 = x_2$ to be the (unique) solution of nonlinear equation $z(1+e^z) = 1000$. The approximate solution of this nonlinear equation is 5.2452. Consequently a KKT solution is

$$\mathbf{x}^* \approx (5.2452; 5.2452),\quad x_0^* = 0.$$

Remark: You can also numerically solve (4) using your favorite solvers (e.g., MATLAB function `fsolve`).

10. (15') Consider standard LP problem

$$\begin{aligned} \text{minimize}_{\mathbf{x}\in R^n} \quad & \mathbf{c}^T\mathbf{x}, \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{LP}$$

with its dual

$$\begin{aligned} \text{maximize}_{\mathbf{y}\in R^m, \mathbf{s}\in R^n} \quad & \mathbf{b}^T\mathbf{y}, \\ \text{subject to} \quad & A^T\mathbf{y} + \mathbf{s} = \mathbf{c}, \quad \mathbf{s} \geq \mathbf{0}. \end{aligned} \tag{LD}$$

For any $\mathbf{x} \in \text{int } \mathcal{F}_p := \{\mathbf{x} \in R^n : A\mathbf{x} = \mathbf{b}, \mathbf{x} > 0\}$ and $\mathbf{s} \in \text{int } \mathcal{F}_d := \{\mathbf{s} \in R^n : \mathbf{s} = \mathbf{c} - A^T\mathbf{y}, \mathbf{s} > \mathbf{0}, \mathbf{y} \in R^m\}$, the **Primal-Dual Potential Function** is defined by

$$\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) := (n+\rho)\log(\mathbf{x}^T\mathbf{s}) - \sum_{j=1}^{n} \log(\mathbf{x}_j \mathbf{s}_j)$$

where $\rho > 0$ is a parameter.

Task: for two LP examples in Problem 5, namely (2) and (3), draw $\mathbf{x}$ part of the primal-dual potential function level sets

$$\psi_6(\mathbf{x}, \mathbf{s}) \leq 0 \quad \text{and} \quad \psi_6(\mathbf{x}, \mathbf{s}) \leq -10,$$

and

$$\psi_{12}(\mathbf{x}, \mathbf{s}) \leq 0 \quad \text{and} \quad \psi_{12}(\mathbf{x}, \mathbf{s}) \leq -10;$$

respectively in int $\mathcal{F}_p$ (on a plane).

Hint: To plot the $\mathbf{x}$ part of the level set of the potential function, say $\psi_6(\mathbf{x}, \mathbf{s}) \leq 0$, you plot

$$\{\mathbf{x} \in \text{int } \mathcal{F}_p : \min_{\mathbf{s}\in\text{int } \mathcal{F}_d} \psi_6(\mathbf{x}, \mathbf{s}) \leq 0\}.$$

This can be approximately done by sampling as follows. You randomly generate N primal points $\{\mathbf{x}^p\}_{p=1}^N$ from int $\mathcal{F}_p$, and N primal points of $\{\mathbf{s}^q\}_{q=1}^N$ from int $\mathcal{F}_d$. For each primal point $\mathbf{x}^p$, you find if it is true that

$$\min_{q=1,\ldots,N} \psi_6(\mathbf{x}^p, \mathbf{s}^q) \leq 0.$$

Then, you plot those $\mathbf{x}^p$ who give an "yes" answer.

Solution: Sample Matlab code:

```

function levelset(n, level, numpoints)

```

```
h = figure;
hold on;

% generate primal feasible solution in the outer 2 loops

for i= 0:1/numpoints:1
    x1 = i;
    for j = 0:1/numpoints:1-x1,
        x2 = j;
        x3 = 1 - x2 - x2;
        % generate dual feasible solution
        for k = 0:-1/numpoints:-15,
            y = k;
            s1 = 1 - y;
            s2 = s1;
            s3 = -y;
           % check level set condition
            if (n * log(x1*s1+x2*s2+x3*s3) - log(x1*x2*x3*s1*s2*s3) < level)
                plot(x1, x2, 'r.');
                break;
            end
        end
    end
end

axis([0 1 0 1]);
%save figure
print(h, '-dpdf', sprintf('n\%ulev\%d.pdf', n, level));
close(h);
```

First, by sampling, it is very hard to plot $\{\psi_6(x,s) \le -10\}$, because here $s = (1+y, 1+y, y) > 0$, so we need $y > 0$. But $\psi_6(x,s) \ge 3\log(x^T s) + 3\log 3 = 3\log(x_1 + x_2 + y) + 3\log 3$. Hence $\{\psi_6(x,s) \le -10\}$ is too harsh for sampled points to survive. Notice that when $n + \rho$ is larger, more primal points survive, and when we look at lower level set $\{\psi \le -10\}$, even though fewer points survive, but they converge to the optimal solution (as we lower the level set again and again).

Here is how we do the analysis: we sample 1000 feasible x in the $\mathcal{F}_p$, which satisfy the conditions $\sum_i x_i = 1, x_i \ge 0$ and for each x, we sample 20 feasible s, where $s = [1-y, 1-y, -y]$, and for s to > 0, we sample $y = -rand(1)$. Then we follow the determine rule in hint, and analyze whether $\min_{q=1,\ldots,N} \psi_6(\mathbf{x}^p, \mathbf{s}^q) \le 0$. or not.

Figure 1: $\psi_6(x, s) \leq 0$, x part

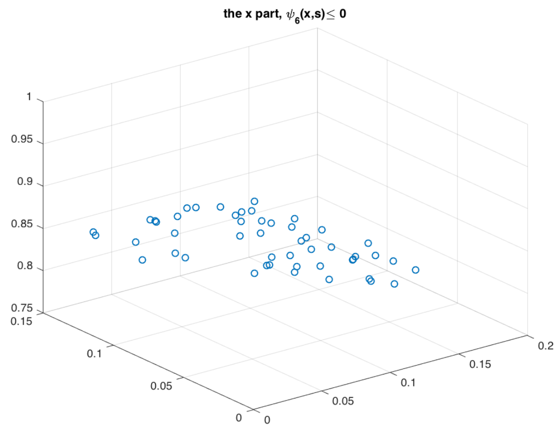


Figure 2: $\psi_{12}(x, s) \leq -10$, x part

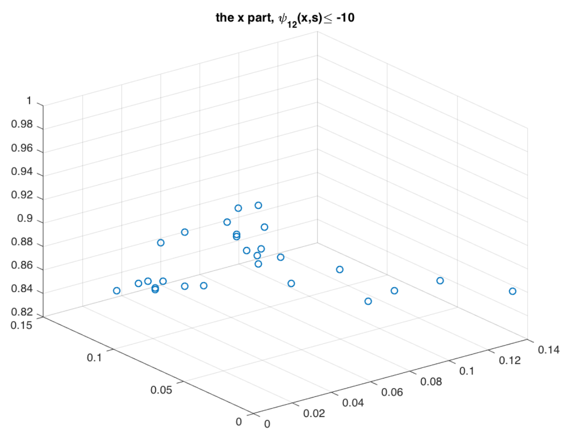
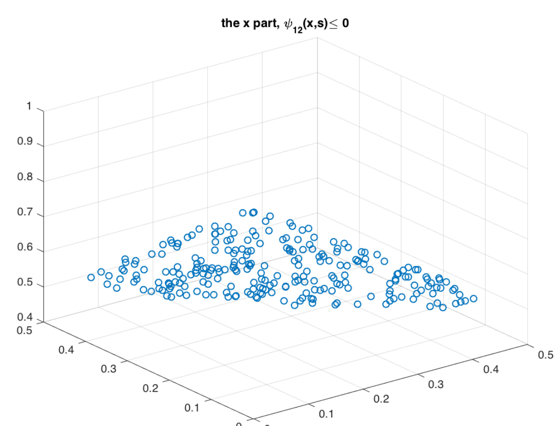


Figure 3: $\psi_{12}(x, s) \leq 0$, x part



Remark. Alternatively, you can use optimization solvers like MOSEK, or fmincon.m in MATLAB, to solve the feasibility problems directly by looping over a grid of x (or uniformly sampling x) and solving the partial feasibility problem in terms of s. If the solver returns infeasibility, then x is not feasible. Otherwise, x is feasible. Similarly, for any sampled/chosen x that needs to be checked, we can simply minimize over s and conclude that x is feasible iff the optimal value of $\psi_{n+\rho}(x,\cdot)$ is non-positive.

11. (10') Recall the Fisher's Equilibrium prices problem (discussed in Lecture Note 6), which we describe here again for reference. Let B be the set of buyers and G be the set of goods. Each buyer $i \in B$ has a budget $w_i > 0$, and utility coefficients $u_{ij} \geq 0$ for each good $j \in G$. Under price $\mathbf{p}$, buyer $i \in B$'s optimal purchase quantity $\mathbf{x}_i^*(\mathbf{p})$ is the solution of the following optimization problem:

$$\begin{aligned} \mathbf{x}_i^*(\mathbf{p}) \in \arg\max \quad & \mathbf{u}_i^T \mathbf{x}_i := \sum_{j \in G} u_{ij} x_{ij} \\ \text{s.t.} \quad & \mathbf{p}^T \mathbf{x}_i := \sum_{j \in G} p_j x_{ij} \leq w_i, \\ & \mathbf{x}_i \geq 0 \end{aligned}$$

Suppose each good $j \in G$ has a supply level $\bar{s}_j$. We call a price vector $\mathbf{p}^*$ an **equilibrium price vector** if the market clears, namely for all $j \in G$,

$$\sum_{i \in B} x^*(\mathbf{p}^*)_{ij} = \bar{s}_j.$$

In the lecture, we discussed how to compute the equilibrium price $\mathbf{p}^*$ and buyers' activities $\{\mathbf{x}_i^*(\mathbf{p}^*)\}_{i \in B}$ under the equilibrium price based on utility coefficients $\{\mathbf{u}_i\}_{i \in B}$, budgets $\{w_i\}_{i \in B}$ and supplies $\bar{\mathbf{s}}$:

$$(\{\mathbf{u}_i\}_{i \in B}, \{w_i\}_{i \in B}, \bar{\mathbf{s}}) \Rightarrow (\mathbf{p}^*, \{\mathbf{x}_i^*(\mathbf{p}^*)\}_{i \in B}) \tag{5}$$

In this question, we consider the inverse problem of (5): suppose the market does not know the "private information" of each buyer, namely the utility $\{\mathbf{u}_i\}_{i \in B}$ and the budgets $\{w_i\}_{i \in B}$, but instead you observe the equilibrium prices $\{\mathbf{p}^{*(k)}\}_{k=1}^K$ and their corresponding realized activities $\{\mathbf{x}_i^{*(k)}\}_{k=1}^K$ under K different supply levels $\bar{\mathbf{s}}^{(1)}, \ldots, \bar{\mathbf{s}}^{(K)}$. The query is to infer buyers' utility coefficients $\{\mathbf{u}_i\}_{i \in B}$ and their budgets $\{w_i\}_{i \in B}$. We assume that the utility function is ℓ_1-normalized, namely $\|\mathbf{u}_i\|_1 = 1$ for $i \in B$.

Hint: Mathematically, the query is to find $\{\mathbf{u}_i\}_{i \in B}$ (s.t. $\mathbf{u}_i \geq \mathbf{0}$ and $\|\mathbf{u}_i\|_1 = 1$) and $\{w_i\}_{i \in B}$ (s.t. $w_i > 0$) such that for all $i \in B$, and $k = 1, \ldots, K$,

$$\begin{aligned} \mathbf{x}_i^{*(k)} = \arg\max_{\mathbf{x}_i} \quad & \mathbf{u}_i^T \mathbf{x}_i \\ \text{s.t.} \quad & (\mathbf{p}^{*(k)})^T \mathbf{x}_i \leq w_i \\ & \mathbf{x}_i \geq \mathbf{0} \end{aligned}$$

given $\{\mathbf{x}_i^{*(k)}\}_{i\in B,k\in\{1,\ldots,K\}}$ and $\{\mathbf{p}^{*(k)}\}_{k\in\{1,\ldots,K\}}$.

Question: Now consider the following 2-buyer 2-good example and solve this inverse problem. Let $B=\{1,2\}$ and $G=\{1,2\}$. Suppose we observe the following 5 scenarios:

- $\mathbf{p}^{*(1)}=(\frac{9}{5};\frac{3}{5})$, $\mathbf{x}_1^{*(1)}=(1;\frac{1}{3})$, $\mathbf{x}_2^{*(1)}=(0;\frac{5}{3})$;
- $\mathbf{p}^{*(2)}=(2;1)$, $\mathbf{x}_1^{*(2)}=(1;0)$, $\mathbf{x}_2^{*(2)}=(0;1)$;
- $\mathbf{p}^{*(3)}=(1;1)$, $\mathbf{x}_1^{*(3)}=(2;0)$, $\mathbf{x}_2^{*(3)}=(0;1)$;
- $\mathbf{p}^{*(4)}=(\frac{1}{2};1)$, $\mathbf{x}_1^{*(4)}=(4;0)$, $\mathbf{x}_2^{*(4)}=(0;1)$;
- $\mathbf{p}^{*(5)}=(\frac{3}{7};\frac{6}{7})$, $\mathbf{x}_1^{*(5)}=(\frac{14}{3};0)$, $\mathbf{x}_2^{*(5)}=(\frac{1}{3};1)$.

Use any approach to find $\{\mathbf{u}_i\}_{i\in B}$ (s.t. $\mathbf{u}_i\geq\mathbf{0}$ and $\|\mathbf{u}_i\|_1=1$) and $\{w_i\}_{i\in B}$ (s.t. $w_i>0$). Describe your approach and report the result.

Solution: Solve the system of KKT conditions: $\mathbf{u}_1=(3/4;1/4)$, $\mathbf{u}_2=(1/3;2/3)$, $w_1=2$, $w_2=1$.