IOE 511 Homework 10
Laura Wendlandt
March 30, 2017

9 Problem 1a.

Consider the following problem (P_c):

$$(\mathrm{P}_c) \quad \min \quad \frac{1}{2}||x-c||_2^2$$
$$\text{s.t.} \quad e^T x = 1$$
$$x \geq 0,$$

where $e \in \mathbb{R}^n$ is a vector of all ones. (The subscript in the problem name refers to the vector c appearing in its objective function.)

Let

$$d = c + \left(\frac{1-e^T c}{n}\right) e$$

(notice that $e^T d = 1$), and consider the problem

$$(\mathrm{P}_d) \quad \min \quad \frac{1}{2}||x-d||_2^2$$
$$\text{s.t.} \quad e^T x = 1$$
$$x \geq 0.$$

Show that (P_c) and (P_d) have the same optimal solutions.

Proof. Both (P_c) and (P_d) have the same feasible region, so in order to show that they have the same optimal solutions, we must show that minimizing their objective functions leads to the same solution. Let's consider the objective function $||x-d||_2^2$ (We'll ignore the $\frac{1}{2}$ because it's also present in the objective function of (P_c)). Using the definition of d, we can expand this function algebraically:

$$||x-d||_2^2 = x^T x - 2x^T d + d^T d = x^x - 2x^T\left(c + \left(\frac{1-e^T c}{n}\right)e\right) + \left(c + \left(\frac{1-e^T c}{n}\right)e\right)^T\left(c + \left(\frac{1-e^T c}{n}\right)e\right)$$

$$= x^T x - 2x^T c + c^T c - 2\left(\frac{1-e^T c}{n}\right)x^T e + 2\left(\frac{1-e^T c}{n}\right)c^T e + \left(\frac{1-e^T c}{n}\right)^2 e^T e.$$

Because of the constraints, we know that $e^T x = 1$. Because of the definition of e, we know that $e^T e = n$. Therefore, we can simplify this further:

$$||x-d||_2^2 = ||x-c||_2^2 - 2\left(\frac{1-e^T c}{n}\right) + 2\left(\frac{1-e^T c}{n}\right)c^T e + \frac{(1-e^T c)^2}{n}$$

$$= ||x-c||_2^2 + \left(\frac{1-e^T c}{n}\right)\left(-2 + 2c^T e + (1-e^T c)\right).$$

We know that $e^T d = e^T c + (1 - e^T c) = 1$. Plugging this in, this simplifies to

$$||x-d||_2^2 = ||x-c||_2^2 + \left(\frac{1-e^T c}{n}\right)\left(-2 + c^T e + e^T d\right) = ||x-c||_2^2 + \left(\frac{1-e^T c}{n}\right)(c^T e - 1).$$

Notice that the second term in this expression does not depend at all on x. Therefore, given a c, minimizing $||x-d||_2^2$ is equivalent to minimizing $||x-c||_2^2$. Therefore (P_c) and (P_d) have the same optimal solutions. $\square$

Problem 1b.

Show that if $d \geq 0$, then $x^\star = d$ is an optimal solution of (P_d).

Proof. Not considering the constraints of (P_d), the objective function $f(x) = \frac{1}{2}||x-d||_2^2$ must be non-negative. Therefore, in the best scenario, the optimal value is zero (and it can never be less than zero). If $d \geq 0$, then d is a feasible point for (P_d). This is because, as noted above, $e^T d = 1$. When $x = d$, $f(x) = \frac{1}{2}||d-d||_2^2 = 0$. Therefore, $x^\star = d$ is an optimal solution of (P_d). □

Problem 1c.

Show that if $d_j < 0$ for some index j, then $x_j^\star = 0$ in any optimal solution of (P_d). (Now use optimality conditions, appropriately justified, along with the fact that $e^T d = 1$.)

Proof. Because all of the constraints are linear, KKT conditions are necessary to characterize an optimal solution $\bar{x}$ for (P_d). The KKT conditions for this problem are as follows:

- *Primal feasibility:* $e^T \bar{x} = 1$ and $\bar{x} \geq 0$.
- *Dual feasibility:* $u \geq 0$.
- *Complementary slackness:* $-u\bar{x} = 0$.
- *Gradient condition:* $\bar{x} - d - ue + ve = 0$.

need more argument

Because x cannot be 0 (because of primal feasibility), $u = 0$ in order to satisfy complementary slackness. □

Problem 1d.

Based on the results proven in parts (a), (b), and (c), state a simple method for solving (P_c) in at most n steps by variable elimination.

1. Given c, calculate d given the formula above.
2. Let I be the set $I = \{0 \leq i \leq n : d_j < 0\}$. For each $i \in I$, set $x_i^\star = 0$.
3. If I is empty, let the remaining (unset) components of $x^\star$ be set to d_i. Stop the algorithm; you've found the optimal solution $x^\star$.
4. Otherwise, for each $i \in I$, remove c_i from the original problem (e.g., if $|I| = 3$ and originally $n = 10$, now c will have three fewer components, and $n = 7$).
5. Go back to step 1.

At each iteration, if I has at least one element, one element of the optimal solution will be set to 0, or if I has no elements, the algorithm will terminate. Therefore, this will take at most n iterations.

Problem 2.

The *support function* of a set $C \subseteq \mathbb{R}^n$ is defined as

$$S_C(y) = \sup\{y^T x : x \in C\}.$$

(We allow $S_C(y)$ to take on value $+\infty$.)

Suppose C and D are closed convex sets in $\mathbb{R}^n$. Show that $C = D$ if and only if their support functions are equal, i.e., for any $y \in \mathbb{R}^n$, $S_C(y) = S_D(y)$.

Proof. Suppose $C = D$. Then $S_C(y) = \sup\{y^T x : x \in C\}$ and $S_D(y) = \sup\{y^T x : x \in D\}$. But, because $C = D$, $S_C(y) = S_D(y)$ by definition. Therefore, if $C = D$, then $S_C(y) = S_D(y)$.

To show the opposite direction, suppose $S_C(y) = S_D(y), \forall y \in \mathbb{R}^n$. Let's assume $C \neq D$. Then either there exists a point $\bar{x} \in D$ such that $\bar{x} \notin C$ or there exists a point $\tilde{x} \in C$ such that $\tilde{x} \notin D$.

If the point $\bar{x}$ exists, then Theorem B.3.1 states that because C is a closed convex set and $\bar{x} \notin C$, there exists a hyperplane $H = \{x : p^T x = \alpha\}$ that strongly separates C and $\bar{x}$. This means that for every $x \in C$, $p^T x < \alpha$, and $p^T \bar{x} > \alpha$. Therefore, $S_C(p) < \alpha$ and $S_D(p) > \alpha$ (because $\bar{x} \in D$), and $S_C(p) \neq S_D(p)$. This is a contradiction. Therefore, if $S_C(y) = S_D(y), \forall y \in \mathbb{R}^n$, then $C = D$.

Using similar logic, if the point $\tilde{x}$ exists, we can construct a hyperplane strongly separating D and $\tilde{x}$ and again reach a contradiction.

Therefore, $C = D$ if and only if their support functions are equal. □

Problem 3.

The purpose of this problem is to show that potentially better Lagrangian dual bounds can be obtained by dualizing as few constraints as possible. Consider the problem

$$\begin{array}{lll} (P) & \min & f(x) \\ & \text{s.t.} & g_i(x) \leq 0, \quad i = 1, \dots, m \\ & & x \in X, \end{array}$$

where X is a subset of $\mathbb{R}^n$. Let r be an integer satisfying $1 \leq r < m$. Consider the set

$$\bar{X} = \{x \in X : g_{r+1}(x) \leq 0, \dots, g_m(x) \leq 0\},$$

and the problem

$$\begin{array}{lll} (\bar{P}) & \min & f(x) \\ & \text{s.t.} & g_i(x) \leq 0, \quad i = 1, \dots, r \\ & & x \in \bar{X}. \end{array}$$

Note that (P) and $(\bar{P})$ are different representations of the same optimization problem. We assume that this problem is feasible and its optimal value is $f^\star > -\infty$.

Let (D) (formed with the help of functions L and $L^\star$) and $(\bar{D})$ (formed with the help of functions $\bar{L}$ and $\bar{L}^\star$) be the Lagrangian duals of problems (P) and $(\bar{P})$, respectively, and $v^\star$ and $\bar{v}^\star$ be the respective optimal values of (D) and $(\bar{D})$.

Show that $v^\star \leq \bar{v}^\star \leq f^\star$.

Proof. First, by weak duality, $\bar{v}^\star \leq f^\star$.

Second, we can formulate the Lagrangians $L(x, u)$ and $\bar{L}(x, u)$ for each optimization problem:

$$L(x, u) = f(x) + \sum_{i=1}^{r} u_i g_i(x) + \sum_{i=r+1}^{m} u_i g_i(x).$$

$$\bar{L}(x, u) = f(x) + \sum_{i=1}^{r} u_i g_i(x).$$

We can also formulate the Lagrangian dual functions $L^\star(u)$ and $\bar{L}^\star(u)$ for each optimization problem:

$$L^\star(u) = \inf_{x \in X} L(x, u).$$

$$\bar{L}^\star(u) = \inf_{x \in \bar{X}} \bar{L}(x, u).$$

Because (P) and $(\bar{P})$ are feasible, there exists a least one point $\tilde{x}$ in the feasible region ($g_i(\tilde{x}) \leq 0, i = 1, \dots, m$). This means that $L^\star(u) \leq \bar{L}^\star(u)$, for $u \geq 0$. This is true because, first, $L^\star(u)$ is minimizing over a larger region than $\bar{L}^\star(u)$ ($\bar{X} \subseteq X$), and, second, for $u \geq 0$, there is at least one x such that the additional term $\sum_{i=r+1}^{m} u_i g_i(x)$ is less than or equal to 0. Thus, the minimum function value of $L(x, u)$ could be smaller than the minimum function value of $\bar{L}(x, u)$ (and at worst, is the same).

Since $L^\star(u) \leq \bar{L}^\star(u)$, $v^\star \leq \bar{v}^\star$. □

Problem 4a.

The performance of a technical device is described by s quality measures $f_j : \mathbb{R}^n \to \mathbb{R}, j = 1, \ldots, s$. The smaller their values, the better the performance is. The requirements of technical feasibility have been formulated in the form of m inequalities:

$$g_i(x) \leq 0, i = 1, \ldots, m.$$

The vector $x \in \mathbb{R}^n$ represents design decisions and $g_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, m$. All functions $f_j(\cdot)$ and $g_i(\cdot)$ are continuously differentiable.

The designer of the device is considering two different approaches to this problem:

Approach 1 Select the most important of the performance measures, say, $f_1(x)$ to minimize, while keeping all the other measures $f_j(x)$ below some target levels $b_j, j = 2, \ldots, s$ selected by the designer.

Approach 2 Create an aggregate objective function

$$f(x) = \sum_{j=1}^{s} w_j f_j(s),$$

where $w_j \geq 0, j = 1, \ldots, s$ are some weights selected by the designer, with $w_1 > 0$.

For each of these two approaches, formulate the corresponding nonlinear optimization problem.

Approach 1

$$\begin{array}{llll} (P_1) & \min & f_1(x) & \\ & \text{s.t.} & f_j(x) \leq b_j, & j = 2, \ldots, s \\ & & g_i(x) \leq 0, & i = 1, \ldots, m \\ & & x \in \mathbb{R}^n. & \end{array}$$

Approach 2

$$\begin{array}{llll} (P_2) & \min & \displaystyle\sum_{j=1}^{s} w_j f_j(x) & \\ & \text{s.t.} & w_1 > 0 & \\ & & w_j \geq 0, & j = 2, \ldots, s \\ & & g_i(x) \leq 0, & i = 1, \ldots, m \\ & & x \in \mathbb{R}^n. & \end{array}$$

Problem 4b.

Prove that for every solution obtained by Approach 2 for some selection of weights $w_j, j = 1, \ldots, s$ with $w_1 > 0$, we can find target levels $b_j, j = 2, \ldots, s$ such that the same solution is optimal in Approach 1.

Proof. Let $w_j, j = 1, \ldots, s$ be a feasible set of weights (ie. $w_1 > 0$ and $w_j \geq 0, j = 1, \ldots, s$). Let $x_2^\star$ be the optimal solution obtained by Approach 2.

Minimizing the objective function $f(x) = \sum_{j=1}^{s} w_j f_j(x)$ is the same as minimizing $f'(x) = f_1(x) + \sum_{j=1}^{s}(w_j/w_1) f_j(x)$ (because $w_1 > 0$, we are allowed to do this division).

Let $b_j = (w_j/w_1) f_j(x_2^\star), b = 2, \ldots, s$. Then, $x_2^\star$ is also an optimal solution for Approach 1. First, we know that $x_2^\star$ is feasible for Approach 1 (because it is in the feasible region for Approach 2). Second, in the best scenario, each $f_j(x) = b_j$. It is not possible to choose an x that does better than this, because if that were possible, $x_2^\star$ would not be the optimal solution for Approach 2. Because each $f_j(x) = b_j$, then the optimal solution must be $x_2^\star$. $\square$

Problem 4c.

Now assume that a constraint qualification is satisfied for both problems. State the first order necessary conditions for optimality for both problems.

Suppose that $\bar{x}$ is an optimal solution for Approach 1. Then the following conditions must be true:

- *Primal feasibility:* $\bar{x}$ is in the feasible region of (P_1).
- *Dual feasibility:* $u \geq 0$.
- *Complementary slackness:*
$$u_j(f_j(x) - b_j) = 0, j = 2, \ldots, s.$$
$$u_{s+i}g_i(x) = 0, i = 1, \ldots, m.$$
- *Gradient condition:*
$$\nabla f_1(x) + \sum_{j=2}^{s} u_j \nabla f_j(x) + \sum_{i=1}^{m} u_{s+i} \nabla g_i(x) = 0.$$

Suppose that $\bar{x}$ is an optimal solution for Approach 2. Then the following conditions must be true:

- *Primal feasibility:* $\bar{x}$ is in the feasible region of (P_2).
- *Dual feasibility:* $u \geq 0$.
- *Complementary slackness:*
$$-u_j w_j = 0, j = 1, \ldots, s.$$
$$u_{s+i}g_i(x) = 0, i = 1, \ldots, m.$$
- *Gradient Condition:*
$$\sum_{i=1}^{s} w_i \nabla f_i(x) + \sum_{i=1}^{m} u_{s+i} \nabla g_i(x) = 0.$$

Problem 4d.

Prove that, if all functions are convex and a constraint qualification is satisfied, for every solution obtained by Approach 1 for some selection of target levels $b_j, j = 2, \ldots, s$, we can find weights $w_j, j = 1, \ldots, s$ with $w_1 > 0$ such that the same solution is optimal in Approach 2.

Proof. Because all of the functions are convex and a constraint qualification is satisfied, KKT conditions are both necessary and sufficient for optimal solutions for both (P_1) and (P_2). Given some selection of target levels $b_j, j = 2, \ldots, s$, let $x^\star$ and $u^\star$ be a KKT point (and thus an optimal solution) to Approach 1. Therefore, we know that $x^\star$ and $u^\star$ satisfy the following conditions:

- $x^\star$ is feasible for Approach 1.
- $u^\star \geq 0$.
- $u_j^\star(f_j(x^\star) - b_j) = 0, j = 2, \ldots, s.$
- $u_{s+i}^\star g_i(x^\star) = 0, i = 1, \ldots, m.$
- $\nabla f_1(x^\star) + \sum_{j=2}^{s} u_j^\star \nabla f_j(x^\star) + \sum_{i=1}^{m} u_{s+i}^\star \nabla g_i(x^\star) = 0.$

Now let $\bar{w}_1 = 1$ and $\bar{w}_j = u_j^\star, j = 2, \ldots, s$. Let $\bar{u}_1 = 0$, let $\bar{u}_j = -(f_j(x^\star) - b_j), j = 2, \ldots, s$, and let $\bar{u}_{s+i} = u_{s+i}^\star, i = 1, \ldots, m$. Then $x^\star$ and $\bar{u}$ are a KKT point (and thus an optimal solution) to Approach 2. We know that this is a KKT point because the following conditions are satisfied:

- $x^\star$ is feasible for Approach 2 (because $x^\star$ is feasible for Approach 1, and the constraints of (P_2) are less restrictive on x than the constraints of (P_1)).

- $\bar{u} \geq 0$. We know that $\bar{u}_1 = 0$, and we know that $\bar{u}_{s+i} \geq 0, i = 1, \ldots, m$ because $u^\star_{s+i} \geq 0, i = 1, \ldots, m$. Further, $f_j(x^\star) \leq b_j$ (because $x^\star$ is feasible for Approach 1), so $\bar{u} = -(f_j(x^\star) - b_j) \geq 0, j = 2, \ldots, s$.
- $-\bar{u}_j \bar{w}_j = 0, j = 1, \ldots, s$. For $j = 1$, this is true because $\bar{u}_1 = 0$. For $j = 2, \ldots, s$, this expression becomes $-\bar{u}_j \bar{w}_j = (f_j(x^\star) - b_j)u^\star_j$, which is one of the KKT conditions for Approach 1 and is equal to 0 because $x^\star$ and $u^\star$ are a KKT point for (P_1).
- $\sum_{i=1}^{s} \bar{w}_i \nabla f_i(x^\star) + \sum_{i=1}^{m} \bar{u}_{s+i} \nabla g_i(x^\star) = 0$. This reduces to $\sum_{i=1}^{s} u^\star_j \nabla f_i(x^\star) + \sum_{i=1}^{m} u^\star_{s+i} \nabla g_i(x^\star) = \nabla f_1(x^\star) + \sum_{j=2}^{s} u^\star_j \nabla f_j(x^\star) + \sum_{i=1}^{m} u^\star_{s+i} \nabla g_i(x^\star) = 0$. This is one of the KKT conditions for Approach 1, and is equal to 0 because $x^\star$ and $u^\star$ are a KKT point for (P_1).

Therefore, for any selection of target levels $b_j, j = 2, \ldots, s$, we can find feasible weights such that the solution to Approach 1 is also the optimal solution to Approach 2. $\square$

Problem 5.

Consider a resistive electric network with node set N and arc set A. Let x_{ij} be the electric current on arc (i, j) and v_i be the voltage of node i. Also, let $R_{ij} > 0$ be the resistance parameters of each arc, and t_{ij} be the (constant) arc voltage (independent of arc current).

The current and voltages in the network must satisfy two basic physical laws:

- Kirchhoff's current law says that for each node i, the total outgoing current is equal to the total incoming current, i.e.,

$$\text{Kirchhoff's law:} \sum_{j:(i,j)\in A} x_{ij} = \sum_{j:(j,i)\in A} x_{ji}, \forall i \in N. \tag{1}$$

- Ohm's law says that the current x_{ij} and the voltage drop $v_i - v_j$ along each arc (i, j) are related by

$$\text{Ohm's law: } v_i - v_j = R_{ij} x_{ij} - t_{ij}, \forall (i, j) \in A. \tag{2}$$

Show that the current x_{ij}, $(i, j) \in A$ that satisfies both laws (1) and (2) is the **unique** solution of the following nonlinear programming problem:

$$\begin{aligned} \min \quad & \sum_{(i,j)\in A} \left(\frac{1}{2} R_{ij} x_{ij}^2 - t_{ij} x_{ij} \right) \\ \text{s.t.} \quad & \sum_{j:(i,j)\in A} x_{ij} = \sum_{j:(j,i)\in A} x_{ij} \qquad \forall i \in N \end{aligned} \tag{3}$$

(note: no nonnegativity constraints!). The objective function of (3) has an energy interpretation; i.e., the current in the system distributes itself so that to minimize the "energy loss" in the system, while satisfying Kirchhoff's current law.

Proof. Because $R_{ij} > 0, \forall i, j$, the function $\frac{1}{2} R_{ij} x_{ij}^2 - t_{ij} x_{ij}$ is convex. Therefore, the objective function of (3) is a sum of convex functions, which is also convex. The constraints are linear, so (3) is a convex optimization problem. Because of this, KKT conditions are both necessary and sufficient. Furthermore, any KKT point (x, u) is a unique global solution.

Before we consider the KKT conditions for this problem, let's consider the gradients of the functions $f(x)$ and $h_k(x), \forall k \in N$:

$$f(x) = \sum_{(i,j)\in A} \left(\frac{1}{2} R_{ij} x_{ij}^2 - t_{ij} x_{ij} \right).$$

$$\frac{\delta f(x)}{x_{ij}} = R_{ij} x_{ij} - t_{ij}, \forall (i, j) \in A.$$

$$h_k(x) = \sum_{j:(k,j)\in A} x_{kj} - \sum_{j:(j,k)\in A} x_{jk}, \forall k \in N.$$

$$\frac{\delta h_k(x)}{\delta x_{ij}} = \begin{cases} 1, & i = k \\ -1, & j = k \\ 0, & \text{otherwise.} \end{cases}$$

Now, let's consider the KKT conditions for this problem. For $\bar{x}$ to be a KKT point (u is irrelevant because there are no inequality constraints), the following two conditions must be met:

- $\bar{x}$ is feasible (e.g. $h_k(\bar{x}) = 0, \forall k \in N$).
- Gradient condition (here, z is used as a multiplier because v represents voltage):

$$\frac{\delta f(\bar{x})}{\delta \bar{x}_{ij}} + \sum_{k \in N} z_k \frac{\delta h_k(\bar{x})}{\delta \bar{x}_{ij}} = R_{ij}\bar{x}_{ij} - t_{ij} + z_i - z_j = 0, \forall (i,j) \in A.$$

Now, let $x_{ij}, (i,j) \in A$ be a current that satisfies both laws (1) and (2) above. Then, we can show that this set of currents also satisfies the KKT conditions laid out above. First, because x satisfies law (1), it is a feasible point for our optimization problem. Second, the gradient condition is satisfied when $z_k = -v_k, \forall k \in N$. Then, the gradient condition becomes:

$$R_{ij}x_{ij} - t_{ij} - v_i + v_j = 0,$$

which can be rewritten as

$$R_{ij}x_{ij} - t_{ij} = v_i - v_j,$$

which is the same as (2) above. Therefore, any current that satisfies laws (1) and (2) also satisfies the KKT conditions. Because KKT conditions are necessary and sufficient and this problem is convex, this current is a unique global solution for our optimization problem. □