EE 159/CS 168 - Convex Optimization
Scott Fullenbaum

Midterm Answers

---

1. (a) S is convex. Let $x_1, x_2 \in S$. Then, for $x_1^T y \leq 1$ and $x_2^T y \leq 1$ with $|y_i| \leq 1$ for all i. To prove S is convex, we need to show that for $\theta \in [0,1]$ that $x_3 = \theta x_1^T y + (1-\theta)x_2^T y \in S$. Since $x_1^T y \leq 1$, then $\theta x_1^T y \leq \theta$. The same can be said for $(1-\theta)x_2^T y \leq (1-\theta)$. Therefore, $\theta x_1^T y + (1-\theta)x_2^T y \leq \theta + 1 - \theta$ so $\theta x_1^T y + (1-\theta)x_2^T y \leq 1$. Therefore, $x_3 \in S$, so S is convex. S is also a polyhedra, as the set where $|y_i| \leq 1$ forms an n-dimensional cube, which is an intersection of half spaces, and $x^T y \leq 1$ is another half space, and a polyhedra is an intersection of a finite number of half spaces and hyperplanes, which the from my descripton, S, is.

   (b) $V = \{x| \; \|x-a\|_2 \leq \theta\|x-b\|_2\} = \{x| \; \|x-a\|_2^2 \leq \theta^2\|x-b\|_2^2\}$
   $= \{x| \;\; (x-a)^T(x-a) \leq \theta^2(x-b)^T(x-b)\}$
   $= \{x| \;\; x^Tx - a^Tx - x^Ta + a^Ta \leq \theta^2(x^Tx - b^Tx - x^Tb + b^Tb)\}$
   $= \{x| \;\; x^Tx - a^Tx - x^Ta + a^Ta - \theta^2(x^Tx - b^Tx - x^Tb + b^Tb) \leq 0\}$
   $= \{x| \;\; (1-\theta^2)x^Tx - a^Tx - x^Ta + a^Ta + \theta^2 b^Tx + \theta^2 bx^T - \theta^2 b^Tb \leq 0\}$
   $= \{x| \;\; (1-\theta^2)x^Tx - (a^Tx - \theta^2 b^Tx) - (x^Ta - \theta^2 x^Tb) + (a^Ta - \theta^2 b^Tb) \leq 0\}$
   $= \{x| \;\; (1-\theta^2)x^Tx - (a-\theta^2 b)^Tx - x^T(a-\theta^2 b) + (a^Ta - \theta^2 b^Tb) \leq 0\}$
   By symmetry of inner product, $(a-\theta^2 b)^Tx = x^T(a-\theta^2 b)$.
   $V = \{x| \;\; (1-\theta^2)x^Tx - 2(a-\theta^2 b)^Tx + (a^Ta - \theta^2 b^Tb) \leq 0\}$
   When $\theta = 1, V = \{x| \; -2(a-b)^Tx + (a^Ta - b^Tb) \leq 0\}$. As there is only one x and this is an inequality, we can see that V is a halfspace which is a convex set.
   Next, for $0 < \theta < 1$, $V = \{x| \;\; (1-\theta^2)x^Tx - 2(a-\theta^2 b)^Tx + (a^Ta - \theta^2 b^Tb) \leq 0\}$. We can notice this is similar to a quadratic form and then factor it as such: First, $V = \{x| \; (1-\theta^2)x^Tx - 2(a-\theta^2 b)^Tx \leq \theta^2 b^Tb - a^Ta\}$ Also, $b^Tb = \|b\|_2^2$, so the inequality becomes:
   $$(1-\theta^2)x^Tx - 2(a-\theta^2 b)^Tx \leq \theta^2\|b\|_2^2 - \|a\|_2^2$$
   $$= x^Tx - \frac{2(a-\theta^2 b)^Tx}{1-\theta^2} \leq \frac{\theta^2\|b\|_2^2 - \|a\|_2^2}{1-\theta^2}$$
   Next, we can complete the square, to start let's add $\|\frac{a-\theta^2 b}{1-\theta^2}\|_2^2$ to both sides. So, the inequality is now:
   $$x^Tx - \frac{2(a-\theta^2 b)^Tx}{1-\theta^2} + \|\frac{a-\theta^2 b}{1-\theta^2}\|_2^2 \leq \frac{\theta^2\|b\|_2^2 - \|a\|_2^2}{1-\theta^2} + \|\frac{a-\theta^2 b}{1-\theta^2}\|_2^2$$
   We can factor this quadratic form and get:
   $$V = \{x|\{(x - \frac{a-\theta^2 b}{1-\theta^2})^T(x - \frac{a-\theta^2 b}{1-\theta^2}) \leq \frac{\theta^2\|b\|_2^2 - \|a\|_2^2}{1-\theta^2} + \|\frac{a-\theta^2 b}{1-\theta^2}\|_2^2\}$$
   From here, we can see that this is a ball centered at $x_0 = \dfrac{a-\theta^2 b}{1-\theta^2}$ with a radius of $r = (\dfrac{\theta^2\|b\|_2^2 - \|a\|_2^2}{1-\theta^2} + \|\dfrac{a-\theta^2 b}{1-\theta^2}\|_2^2)^{1/2}$. As this is a convex ball, it is a convex set.

2. (a) $f(x, y)$ is convex and quasiconvex. To show, $f(x, y)$ is actually a perspective function of the norm squared of an affine transformation. Consider the function $h(w, y) = w^T w/y$ where $w \in \mathbb{R}^n$ and $y \in \mathbb{R}$. This is the perspective of $w^T w$ and is stated to be convex in Example 3.1 (Book page 89). We can rewrite $w^T w = \|w\|_2^2$. So $h(w, y) = \|w\|_2^2/y$. $Ax-b$ is an affine mapping as $Ax - b = Ax + (-b)$. So, we can see that $h(Ax - b, y) = f(x, y)$. $h(Ax - b, y)$ is the composition of a convex function with an affine mapping so it is convex. Therefore, $f(x, y)$ is convex. Since $f(x, y)$ is convex, it is also quasiconvex.

(b) $f(x)$ is convex and quasiconvex. To prove, $\mathbf{epi}f = \{(x, t) \in \mathbb{R}^{n+1} \mid \|x\|_2 < 1, \dfrac{\|Ax - b\|_2^2}{1 - x^T x} \leq t\}$.
Since $x^T x = \|x\|_2^2$, we can rewrite it as $\mathbf{epi}f = \{(x, t) \in \mathbb{R}^{n+1} \mid \|x\|_2 < 1, \dfrac{\|Ax - b\|_2^2}{1 - \|x\|_2^2} \leq t\}$.
Looking at $\dfrac{\|Ax - b\|_2^2}{1 - \|x\|_2^2} \leq t \to \|Ax - b\|_2^2 \leq t(1 - \|x\|_2^2) \leq t$.
$\mathbf{epi}f = \{(x, t) \in \mathbb{R}^{n+1} | \ \|x\|_2^2 < 1, \|Ax - b\|_2^2 \leq t\}$. Therefore, $\mathbf{epi}f$ is an affine function applied to a second order cone. A second order cone is convex (Book page 30), and affine functions preserve convex sets, so $\mathbf{epi}f$ is convex. As we know a function is convex if and only if it's epigraph is convex. Since the epigraph is convex, then the function is convex. Additionally, as the function is convex that implies it is also quasiconvex.

3. (a) First, $x^T \beta + v = \langle (1, x), (v, \beta) \rangle$ where $\beta \in \mathbb{R}^n$. We are adding a 1 and v to the dot product and that allows us to merge it into one term. Let the column vector $\hat{\beta} = [v, \beta_1, \beta_2, ..., \beta_n]^T$. It is clear our model becomes $\hat{y} = x^T \hat{\beta}$ for any vector x in our set. As, there are m measurements in $\hat{y}$, we can create a linear system of m equations, represented in the following form:
$\begin{bmatrix} 1 & ..., & x_1^T & ... \\ 1 & ... & x_2^T & ... \\ ... & ... & ... & ... \\ 1 & ..., & x_m^T & ... \end{bmatrix} \begin{bmatrix} v \\ \beta_1 \\ ... \\ \beta_n \end{bmatrix} = \begin{bmatrix} \hat{y_1} \\ \hat{y_2} \\ ... \\ \hat{y_m} \end{bmatrix}$ I will express this as $X\hat{\beta} = \hat{y}$
Though I wrote this as an equality, we cannot guarantee this equation has a solution. The goal of the least squares problem is to make $X\hat{\beta}$ as close to $\hat{y}$ as possible. Stated compactly, and as X and $\hat{y}$ are known, the least squares problem finds the $\hat{\beta}$ that minimizes $f(\hat{\beta}) = \|X\hat{\beta} - \hat{y}\|_2^2$. $f(\hat{\beta})$ is a convex function, as $f(\hat{\beta})$ is taking the composition of the affine mapping $X\hat{\beta} - \hat{y}$ with the L_2 norm squared. The L_2 norm squared is a convex function, and affine mappings preserve convexity, so $f(\hat{\beta})$ is convex, and as we are minimizing a convex function, this is a convex problem. Additionally, the exact solution is $\hat{\beta} = X^t \hat{y}$ where X^t is the psuedoinverse of X (Book page 153).
For the second part, it is a bad to idea to include the mean feature. The matrix X becomes: $\begin{bmatrix} 1 & ..., & x_1^T & ... & \tilde{x_1} \\ 1 & ... & x_2^T & ... & \tilde{x_2} \\ ... & ... & ... & ... & ... \\ 1 & ..., & x_m^T & ... & \tilde{x_m} \end{bmatrix}$
X now has dimension $m \ x \ (n + 2)$, so to calculate $X\hat{\beta}$, $\beta \in \mathbb{R}^{n+2}$ which means adding another parameter and could result in overfitting and bias results towards the mean.
To expand on why this doesn't improve results, though this wasn't discussed in class, due to the Best Approximation theorem, the optimal solution to the least squares problem is

equivalent to taking the orthogonal projection onto the subspace formed by the columns of X. However, the mean feature is clearly a linear combination of the other columns in X, so it isn't be independent and is already in the subspace. Therefore, including the mean vector won't actually improve results, as we're introducing a feature that's already accounted for, and least squares is already optimal.

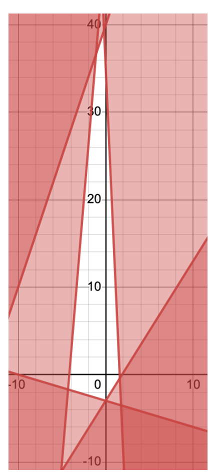
(b) As the problem states to maximize profit under a total cost constraint, I denote the maximum cost as C_{max}. I put an inequality constraint on total cost, as the company would probably prefer if they could maximize profit without using their whole budget. Additionally, from the problem statement we can see that $v = Rc$. This leads to the following problem:

$$\begin{gathered} \text{maximize } a^T v \\ \text{subject to: } Rc = v \\ \mathbf{1}^T c \leq C_{max} \end{gathered}$$

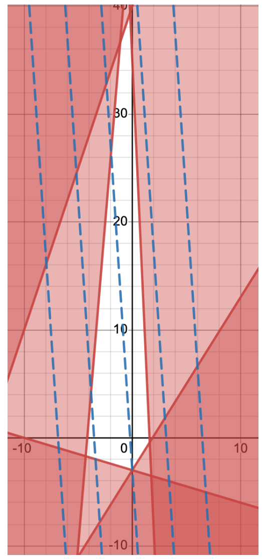
This is a linear program and a convex problem as there are linear constraints and a convex objective function. To change into the general form of a linear program, the objective function is now minimizing $-a^T v$. So, the resulting linear program is:

$$\begin{gathered} \text{minimize } -a^T v \\ \text{subject to: } Rc = v \\ \mathbf{1}^T c \leq C_{max} \end{gathered}$$

4. (a) To start, I consider the x-axis to be x_1 and y-axis as x_2. The system of constraints of $Ax \leq b$ can be expressed as the following system of inequalities:
$-6x_1 - 20x_2 \leq 60$
$22x_1 + x_2 \leq 35$
$-24x_1 + 2x_2 \leq 100$
$-3x_1 + x_2 \leq 40$
$8x_1 - 5x_2 \leq 15$
I made the graphs in Desmos, and they can be viewed here:
https://www.desmos.com/calculator/h4xthz1zqf
The constraints lead to this graph of the feasible set $\mathcal{F}$, which is the region in white. $\mathcal{F}$ is the intersection of linear inequalities which are half spaces and are convex, so $\mathcal{F}$ is convex:



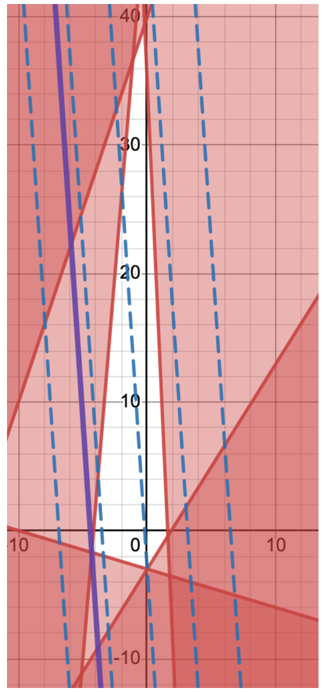
(b) Each level set in L_α is a dotted blue line, with L_{-100} being the one furthest on the left, and them moving to the right for each successive step:



Only the level sets at $\alpha = -50$ and $\alpha = 0$ pass through $\mathcal{F}$. However, no point on either of these level sets is optimal. To verify, you can check that no point on either line satisfies the first order condition for optimality. I'm only showing for $\alpha = -50$, but the process is identical. Consider (x_1, x_2) on the level set for $\alpha = -50$, so $15x_1 + x_2 + 3 = -50$ or $15x_1 + x_2 = -53$. Checking the first order condition for optimality, and let $(y_1, y_2) \in \mathcal{F}$, then $\nabla f(x)^T(y-x) = [15, 1]([y_1, y_2]^T - [x_1, x_2]^T)$. Simplifying this becomes $15y_1 + y_2 - (15x_1 + x_2) = 15y_1 + y_2 + 53 \geq 0$. If this holds for all $y \in \mathcal{F}$, then (x_1, x_2) is optimal. Visually, this is the region to the right of the level curve at $\alpha = -50$, which clearly doesn't include all of $\mathcal{F}$. This process can be repeated for $\alpha = 0$. Also, as solutions to convex problems are unique, and this is a convex problem, if an arbitrary point on either level set in $\mathcal{F}$ was optimal, that would imply infinitely many solutions, which cannot be true.

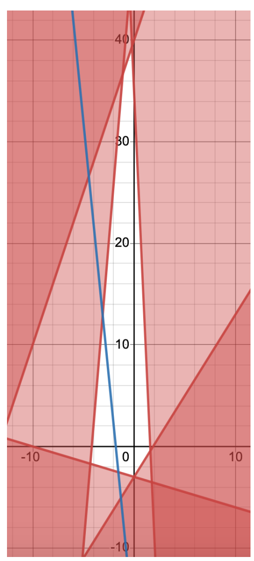
Visually, the true optimal solution is at the left most corner point. The reasoning behind this is as α decreases, the level set will shift left until it is no longer in $\mathcal{F}$. However, we can pick an α such that the objective function only intersects with that corner which is a single point in $\mathcal{F}$, which means it's also unique. and will also be the minimum, because if α decreases any further, then $L_\alpha \nsubseteq \mathcal{F}$.

As my answers are in decimals, there might be some small rounding error. The left-most corner point is the intersection of the lines $24x_1 + 2x_2 = 100$ and $-6x_1 - 20x_2 = 60$, which is at the point $x^* = (-4.309, -1.707)$. Using the objective function, L_{p^*} is the level set passing through this point, which is $15x_1 + x_2 + 3 = -63.342$. This is visualized on the plot below, and L_{p^*} is the thick purple line:



Visually, the point $x^* = (-4.309, -1.707)$ satisfies the optimality condition, so x^* is the optimal point associated with the unique minimum value $p^* = -66.342$.

(c) The new constraint $g^T x = h$, is the equation $20x_1 + 2x_2 = -36$. The new feasible set, $\mathcal{F}_2$ is the intersection between the old feasible set and this line, which is a line segment. In this picture, $\mathcal{F}_2$ is the blue line in the old feasible set:



To solve, we want to find the minimum α such that L_α and our constraint intersect in $\mathcal{F}_2$. Visually, it appears that the left most point (x_1, x_2), where $g^T x = h \in \mathcal{F}_2$ is the optimal solution. This is where the lines $-24x_1 + 2x_2 = 100$ and $20x_1 + 2x_2 = -36$ intersect, which is at the point $x^* = (-3.091, 12.91)$, plugging this into the objective function gives $p^* = -30.456$. To show this is optimal, we can check the optimality

condition, $\nabla f(x^*)^T(y - x^*) \geq 0$ for $y \in \mathcal{F}_2$. This equation becomes $[15, 1]([y_1, y_2]^T - [-3.091, 12.909]^T) \to 15y_1 + y_2 \geq -33.456$. To verify we satisfy this constraint on $\mathcal{F}_2$, it is first apparent that we satisfy this equation at $(y_1, y_2) = x^*$, which is the left most point on $\mathcal{F}_\in$. We can also pick the furthest point from the optimal solution on $\mathcal{F}_2$ which can be graphically determined to be $(-1.546, -2.536)$. We satisfy the above optimality constraint at this point. Since the constraint is a line segment, and we have two points that satisfy the condition, then it follows that every point on the line segment between $((-3.091, 12.91)$ and $(-1.546, -2.536)$, or $\mathcal{F}_2$ will also satisfy the optimality condition.
I'm not sure how much theory you want, but the idea behind this statement is if you pick one point on the constraint (x_1, x_2) and one on the level set (y_1, y_2), such that $x_2 = y_2$, the distance between them is 0 only when the lines intersect, and $x_1 - y_1$ increases as x_2 decreases. So the distance between the level set and any point on the constraint is ≥ 0, and all points on the line segment satisfy the optimality condition.
In conclusion the optimal solution to this problem with the added constraint is $x^* = (-3.091, 12.91)$ associated with the unique minimum of $p^* = -30.456$.

5. (a) As r, the vector of materials is chosen ahead of time, the cost is fixed beforehand. Quantity produced, q, is a vector representing how much of each product is produced and is set before we know demand. There are K demand vectors, and the equation for revenue, and for any demand vector, $d^{(i)}$, as stated in the problem, the amount of revenue associated with this is $p^T min\{q, d^{(i)}\}$. The probability associated with the actual demand being each potential demand vector is π_i, meaning the expected revenue can be written as $\sum_{i=1}^{K} \pi_i p^T min\{q, d^{(i)}\}$. In addition to the provided constraint that $Aq \preceq r$, a non-negative amount of materials must be used and a non-negative amount of products used be produced, so $0 \preceq r$ and $0 \preceq q$. This leads to the following problem:

$$\begin{aligned} \text{maximize } & \sum_{i=1}^{K} \pi_i p^T min\{q, d^{(i)}\} - c^T r \\ \text{subject to: } & Aq \preceq r \\ & 0 \preceq r \\ & 0 \preceq q \end{aligned}$$

This can be rewritten as a convex problem like below. The objective function is convex and all constraints are linear and therefore convex so this is a convex problem.

$$\begin{aligned} \text{minimize } & c^T r - \sum_{i=1}^{K} \pi_i p^T min\{q, d^{(i)}\} \\ \text{subject to: } & Aq \preceq r \\ & 0 \preceq r \\ & 0 \preceq q \end{aligned}$$

(b) Since r is still chosen ahead of time, $c^T r$ is still fixed. Since d is now known before q is decided, we can introduce K production vectors, $q^{(1)}, q^{(2)}, ..., q^{(K)}$ associated with each $d^{(i)}$. Furthermore, there is no reason to produce more than the demand as it won't sell, so, we can introduce a new constraint that $0 \preceq q^{(i)} \preceq d^{(i)}$ for all i. Additionally, $0 \preceq q^{(i)}$ as we must produce a non-negative amount by problem assumptions. As a result, the min function in the previous objective function is replaced by $q^{(i)}$ as under this new constraint, the min will always be $q^{(i)}$. We also need additional constraints to account for each $q^{(i)}$ and $d^{(i)}$ as q is not set beforehand and we have now introduced a constraint that relies on demand. The old constraint on r remains. The new problem is:

$$\begin{gathered}\text{maximize} \sum_{i=1}^{K} \pi_i p^T q^{(i)} - c^T r \\ \text{subject to: } Aq^{(i)} \preceq r \\ 0 \preceq r \\ 0 \preceq q^{(i)} \preceq d^{(i)} \\ \text{for } i = 1, \ldots.K\end{gathered}$$

This can be rewritten as:

$$\begin{gathered}\text{minimize } c^T r - \sum_{i=1}^{K} \pi_i p^T q^{(i)} \\ \text{subject to: } Aq^{(i)} \preceq r \\ 0 \preceq r \\ 0 \preceq q^{(i)} \preceq d^(i) \\ \text{for } i = 1, \ldots.K\end{gathered}$$

We have convex constraints and have a convex objective function, so this is a convex problem.
I believe that the second methodology will generally lead to higher profits. Due to it being a system of many constraints, I hesitate to definitively conclude from the objective function and constraints that this will always hold. However, as cost is fixed for both, we need to only look at what better maximizes revenue. For the first formulation, if $q^{(i)}$ is produced, the demand is still restricting revenue. So, we will be limiting our revenue as long as $d \preceq q$. However, in method two, we account for this with the added constraint, so we won't have leftover products and can produce more/less of specific goods which means that in general should be associated with higher expected revenue.

8. (a) If $f(x)$ is convex, then $\nabla^2 f(x) \succeq 0$. To find $\nabla^2 f(x)$, we can use the multivariate chain rule, where $f(x) = h(g(x))$ and $h(y) = -exp(-y)$ and $f(x) : \mathbb{R}^n \to \mathbb{R}$:
$\nabla^2 f(x) = -exp(-g(x))\nabla g(x)\nabla g(x)^T + exp(-g(x))\nabla^2 g(x)$
$\nabla^2 f(x) = exp(-g(x))(\nabla^2 g(x) - \nabla g(x)\nabla g(x)^T)$
$exp(-g(x)) > 0 \ \forall x \in \mathbb{R}^n$, so to check if $\nabla^2 f(x) \succeq 0$, we will need to verify that $\nabla^2 g(x) - \nabla g(x) * 1 * \nabla g(x)^T \succeq 0$ Then, we can construct a block matrix that contains all terms X = $\begin{bmatrix} \nabla^2 g(x) & \nabla g(x) \\ \nabla g(x)^T & 1 \end{bmatrix}$
X is symmetric as the Hessian matrix is always symmetric, and $\nabla g(x)$ transposed ends up where $\nabla g(x)^T$ is and 1 remains where it already was. Since X is symmetric, I believe we can use properties of the Schur complement to finish the problem. Following these means $X \succeq 0$ if and only if $\nabla^2 g(x) \succeq 0$, and as $g(x)$ is convex this holds. Furthermore, $X \succeq 0$ if and only if $\nabla^2 g(x) - \nabla g(x)\nabla g(x)^T \succeq 0$. This means $f(x)$ is convex if $X \succeq 0$, which is what we want to prove. The properties of the Schur Complement used are shown in the book on pages 650-651.

(b) P is a permutation matrix, so it just permutes the rows and columns of A. Therefore, $APx - b$ is an affine transformation. The norm function is also convex, and the composition of a convex function with an affine transformation is also convex. So, $\|APx - b\|$ is convex. Additionally, taking the max of a set of convex functions is convex, (Book page 80) therefore $f(x)$ is convex.