ACM 113 Homework 2

October 31, 2007

Problem 1

The inequality defining convexity relates the value of a convex function for 3 points along an arbitrary line. That is, the points x, y and $\theta x + (1-\theta)y$ are collinear by construction, and convexity is just the relation $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$ for $0 \leq \theta \leq 1$. Conversely any 3 collinear points can be expressed in the form $\{x, y, \theta x + (1-\theta)y\}$ with $0 \leq \theta \leq 1$. This implies a function is convex if and only if its restriction along an arbitrary line does so too.

(a) There are a few different ways to manipulate the expression for the trace(X^{-1}) to show it is convex. What follows is just one of them. Noting that the trace of X^{-1} is the sum of the n largest eigenvalues of X^{-1}, we have:

$$f(X) = \max_{y_i^T y_j = \delta_{ij}} \sum_{i=1}^{n} y_i^T X^{-1} y_i$$

We want to show that $y^T X^{-1} y$ is a convex function of X. This will imply $\sum_{i=1}^{n} y_i^T X^{-1} y_i$ is a sum a convex functions, hence convex. In turn, this will imply that f is the maximum of convex functions, hence convex. Consider the function $y^T X^{-1} y$ along a line $X = Z + tV$. Then:

$$g(t) = y^T (Z + tV)^{-1} y = \tilde{y}^T (I + tW)^{-1} \tilde{y}$$

Where $W = Z^{-1/2} V Z^{-1/2}$ and $\tilde{y} = Z^{-1/2} y$. For $t < \|W\|$, $(I + tW)^{-1}$ has power series expansion $\sum_{n=0}^{\infty} (-1)^n t^n W^n$, where the infinite sum is defined as the limit of partial sums under a matrix norm. This justifies the following derivatives:

$$g'(t) = -\tilde{y}^T (I + tW)^{-2} W \tilde{y}$$

$$g''(t) = 2\tilde{y}^T W (I + tW)^{-3} W \tilde{y} = 2\hat{y}^T (Z + tV)^{-1} \hat{y}$$

Where $\hat{y} = Z^{1/2}(I + tW)^{-1} W Z^{-1/2} y$. Since $Z + tV > 0$ we have $g''(t) \geq 0$. So g is a convex function of t, which means $y^T X^{-1} y$ is a convex function of X and we are done.

(b) We have $f(X) = (\det X)^{1/n}$ and consider the function along an arbitrary line $X = Z + tV$,

$$g(t) = \det(Z + tV)^{1/n} = \det(Z)^{1/n} \det(I + tW)^{1/n} = \det(Z)^{1/n} \prod_{i=1}^{n} (1 + t\lambda_i)^{1/n}$$

Where $W = Z^{-1/2} V Z^{-1/2}$ and λ_i are the eigenvalues of W. So we compute the derivative:

$$g'(t) = g(t) \frac{1}{n} \sum_{i=1}^{n} \frac{\lambda_i}{1 + t\lambda_i}$$

Noting that $g' = gh \Rightarrow g'' = g(h' + h^2)$.

$$g''(t) = g(t) \left(-\frac{1}{n} \sum_{i=1}^{n} \left(\frac{\lambda_i}{1 + t\lambda_i} \right)^2 + \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\lambda_i}{1 + t\lambda_i} \right)^2 \right)$$

But for an arbitrary vector $a = (a_1, \ldots, a_n)$ we have

$$-n\sum_{i=1}^{n} a_i^2 + \left(\sum_{i=1}^{n} a_i\right)^2 = -(1^T 1)(a^T a) + (a^T 1)^2 \leq 0$$

by the Cauchy-Schwartz inequality. So, letting $a_i = \frac{\lambda_i}{1+t\lambda_i}$, in the above inequality, and noting $g(t) \geq 0$ since $Z + tV \in S_n^{++}$, we get $g''(t) \leq 0$. Hence g is concave, likewise f is concave.

Problem 3.34

(a) Given nonzero x, consider the ray generated by x, or the set R $= \{sx : s > 0\}$. We have 3 cases to consider.
The ray does not hit C. That is, $R \cap C = \emptyset$, so $M_C(x) = \inf \emptyset = +\infty$ by convention.
The ray hits C and passes through it. That is, the intersection $R \cap C$ is nonempty and bounded, so if $s^* = \sup\{s : sx \in C\}$, we have $M_C(x) = 1/s^*$.
The ray hits C and stays in it. That is, the intersection $R \cap C$ is nonempty and unbounded, so $M_C(x) = 0$.
Finally, note if $0 \in C$, then $M_C(0) = \inf\{t : t > 0\} = 0$. If $0 \notin C$, we take the convention $M_C(0) = 0$ to enforce homogeneity.
(b) First, for $\alpha = 0$, we have $0M_C(x) = M_C(0x)$ as shown in part (a). Otherwise for nonzero α, we need 2 properties of $M_C(x)$ which are obvious from the definition:

$$M_C(x)^{-1}x \in \bar{C}$$

$$t^{-1}x \in \bar{C} \Rightarrow M_C(x) \leq t$$

(Here we denote the closure of C as $\bar{C}$.) So for any $\alpha > 0$, the first property implies $(\alpha M_C(x))^{-1}(\alpha x) \in \bar{C}$, which implies $M_C(\alpha x) \leq \alpha M_C(x)$ by the second property. Similarly,

$$(M_C(\alpha x))^{-1}(\alpha x) \in \bar{C} \Rightarrow (\alpha^{-1} M_C(\alpha x))^{-1} x \in \bar{C} \Rightarrow M_C(x) \leq \alpha^{-1} M_C(\alpha x)$$

Together these inequalities imply $M_C(\alpha x) \leq \alpha M_C(x) \leq M_C(\alpha x)$, or $M_C(\alpha x) = \alpha M_C(x)$
(c) The domain is the set of points for which $t^{-1}x \in C$ for some $t > 0$. This is the conic hull of C.
(d) By part (c), the domain of M_C is convex. For any x, y in the domain of M_C, and $\theta \in [0, 1]$, we want to show $M_C(\theta x + (1-\theta)y) \leq \theta M_C(x) + (1-\theta)M_C(y)$, or equivalently by definition of M_C,

$$\frac{\theta x + (1-\theta)y}{\theta M_C(x) + (1-\theta)M_C(y)} \in C$$

For $t_x, t_y > 0$ with $x/t_x, y/t_y \in C$, we have

$$\frac{\theta x + (1-\theta)y}{\theta t_x + (1-\theta)t_y} = \frac{\theta t_x (x/t_x) + (1-\theta)t_y (y/t_y)}{\theta t_x + (1-\theta)t_y} \in C$$

Where we use the convexity of C and the fact that $\theta t_x/(\theta t_x + (1-\theta)t_y) \in [0, 1]$. Therefore, for all t_x, t_y as described above,

$$M_C(\theta x + (1-\theta)y) \leq \theta t_x + (1-\theta)t_y$$

Taking the infimum over all such t_x, t_y gives the convexity relation as desired.
(e) Obviously, $M_C \geq 0$. Since C is symmetric, we have $M_C(-x) = M_C(x)$. And M_C is homogeneous, we have

$$M_C(\alpha x) = M_C(|\alpha| x) = |\alpha| M_C(x)$$

Also since C is bounded, $M_C(x) = 0$ only if $x = 0$. Since M_C is convex and homogeneous, it satisfies the triangle inequality:

$$M_C\left(\frac{1}{2}x + \frac{1}{2}y\right) \leq \frac{1}{2}M_C(x) + \frac{1}{2}M_C(y) = M_C\left(\frac{1}{2}x\right) + M_C\left(\frac{1}{2}y\right)$$

The unit ball in this norm is C itself.

Problem 4.3

We compute the eigenvalues of P to be $0.259, 13.84, 27.89$, so P is positive definite, so the function is convex. We compute $\nabla f(x^*) = Px^* + q = (-1, 0, 2)$. Note for all feasible y, we have $(y_1 - x_1^*) \leq 0$ and $(y_3 - x_3^*) \geq 0$. So for all such y we have:

$$\nabla f(x^*)^T(y - x^*) = (-1)(y_1 - x_1^*) + 2(y_3 - x_3^*) \geq 0$$

So we conclude x^* satisfies the optimality criterion for convex differentiable functions.

Problem 4.8

(a) Let $A^\dagger$ be the Moore-Penrose inverse of A. If $AA^\dagger b \neq b$, then the feasible set is empty.
Otherwise, let $P = I - A^\dagger A$ be the projection operator onto the null space of A. If $Pc \neq 0$, then there is no solution because the function is unbounded. That is, $x(\theta) = A^\dagger b - \theta Pc$ is feasible, but $c^T x(\theta) \to -\infty$ as $\theta \to \infty$.
If, however, $Pc = 0$, then optimal solutions exist, but all feasible solutions are equally good. So we can choose the one given by the Moore-Penrose inverse: $x^* = A^\dagger b$.
(b) If $c = \lambda a$ for $\lambda \leq 0$, then $x^* = \frac{ab}{a^T a}$ is a solution. Otherwise there is no solution because the function is unbounded. That is, let $x(\theta) = -\theta\left(\frac{a}{a^T a} + \frac{c}{c^T c}\right)$. Then as $\theta \to \infty$, $x(\theta)$ is feasible, and $c^T x(\theta) \to -\infty$.
(c) The objective is a sum of term $c_i x_i$, and each constraint depends only on one variable. Therefore we can solve by minimizing each component of x independently. So one optimal solution is given by $x_i^* = l_i$ if $c_i \geq 0$ and $x_i^* = u_i$ if $c_i < 0$
(d) Again, without loss of generality, sort the indices, so that $c_1 \leq c_2 \ldots \leq c_n$. So for all feasible x, $c^T x \geq c_1$. So one optimal solution is just e_1.
If the equality is replaced with an inequality constraint, if $c_1 \leq 0$ we still have $c^T x \geq c_1$ for all feasible x, so the answer will be the same. If however, $c_1 > 0$, then we have, $c^T x \geq 0$ for all feasible x, in which case $x^* = 0$ is optimal.
(e) Without loss of generality, sort the indices, so that $c_1 \leq c_2 \ldots \leq c_n$. Then we have $x_i^* = 1$ for $i \leq \alpha$, and $x_{\lfloor \alpha \rfloor + 1}^* = \alpha - \lfloor \alpha \rfloor$, and $x_i^* = 0$ for $i \geq \alpha + 1$. It is straightforward to see that $c^T x \geq c^T x^*$ for all feasible x.
If we replace the equality with an inequality constraint, then the answer is the same, except $x_i = 0$ when $c_i > 0$.
(f) Now we sort the indices, so that $c_1/d_1 \leq c_2/d_2 \ldots \leq c_n/d_n$. Then define k and β to satisfy $0 \leq \beta \leq 1$ and

$$\sum_{i=1}^{k-1} d_i + \beta d_k = \alpha$$

So then $x_i^* = 1$ for $i < k$ and $x_k^* = \beta$, and $x_i^* = 0$ for $i > k$.
To see that this is optimal, we can make a change of variables $y_i = d_i x_i$ transforming the problem to

$$\min \sum_{i=1}^{n} (c_i/d_i) y_i$$

$$\text{s.t.: } 1^T y = \alpha, 0 \leq y \leq d$$

In this setting it is clear that if $y_i^* = d_i x_i^*$, then $\sum (c_i/d_i) y_i \geq \sum (c_i/d_i) y_i^*$ for all feasible y.

Problem 4.12

At node i, the flow out minus the flow in must equal b_i (the net supply or demand). We just need to write out as an equation. Then the constraints and objective are obviously linear, so the problem need not be recast.

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

$$\text{s.t.: } \sum_{j=1}^{n} x_{ij} - x_{ji} = b_j, \text{ for } i = 1 \dots n$$

$$l_{ij} \leq x_{ij} \leq u_{ij} \text{ for } i, j = 1 \dots n$$

Problem 4.25

If we found such an a b, we would have for $i = 1, \dots, K$:

$$\inf_{x \in E_i} a^T x + b = \inf_{\|u\|_2 \leq 1} a^T (P_i u + q_i) + b > 0$$

$$\sup_{\|u\|_2 \leq 1} -a^T (P_i u + q_i) - b < 0 \Leftrightarrow \sup_{\|u\|_2 \leq 1} -u^T P_i a = \|P_i a\|_2 < q_i^T a + b$$

Similarly, for $i = K+1, \dots, K+L$

$$\sup_{\|u\| \leq 1} a^T (P_i u + q_i) + b < 0 \Leftrightarrow \|P_i a\|_2 < -q_i^T a - b$$

In summary, we have $\|P_i a\|_2 < y_i(q_i^T a + b)$, where $y_i = 1$ for $i = 1, \dots, K$, and $y_i = -1$ for $i = K+1, \dots, K+L$. Note that given a solution, we can scale (a, b) by a positive constant without changing the hyperplane. And since these are strict inequalities, we can scale so that the difference between the two sides is at least 1 for every i. That is, the condition becomes $\|P_i a\|_2 \leq y_i(q_i^T a + b) - 1$.
Remark: The problem only calls for a feasibility condition, but while we're at it, we can try to minimize the norm of a, so that at least one of these inequalities is exact. In summary, we have the SOCP:

$$\min t$$

$$\|P_i a\|_2 \leq q_i^T a + b - 1 \text{ for } i = 1 \dots K$$

$$\|P_i a\|_2 \leq -q_i^T a - b - 1 \text{ for } i = K+1 \dots K+L$$

$$\|a\|_2 \leq t$$

Any feasible solution (a, b, t) will give a separating hyperplane. Conversely, any pair (a, b) which corresponds to a separating hyperplane can be scaled appropriately to give a feasible solution to the SOCP. In fact, one can show that the optimal solution maximizes the distance from the hyperplane to the closest ellipsoids, which is equal to $1/t$ for the optimal solution.