SELECTED SOLUTIONS, SECTION 2.3

1. Prove by induction that if the functions g_0, g_1,…,$g_m : \mathbb{E} \to \mathbb{R}$ are all continuous at the point $\bar{x}$ then so is the max-function $g(x) = \max_i\{g_i(x)\}$.

We first note that the function $\max : \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto \max\{x, y\}$ is continuous. (For those who want details here: If (a, b) is an open interval, then

$$\max^{-1}(a,b) = \big\{(x,y) \in \mathbb{R}^2 : x < b \text{ and } y < b\big\} \cap \big\{(x,y) \in \mathbb{R}^2 : x > a \text{ or } y > a\big\},$$

which is the intersection of two open sets and thus open.) As a consequence, if g_1, $g_2 : \mathbb{E} \to \mathbb{R}$ are continuous at $\bar{x}$, then so is $\max\{g_1, g_2\}$, because it is the composition of the continuous function max with the function $x \mapsto (g_1(x), g_2(x))$, which by assumption is continuous at $\bar{x}$.

Now assume that we have shown that the maximum of any n functions that are continuous at $\bar{x}$ is continuous at $\bar{x}$ as well. Now assume that g_1,…,g_{n+1} are all continuous at $\bar{x}$. Then we can write

$$g(x) := \max\{g_1(x), \ldots, g_{n+1}(x)\} = \max\{\max\{g_1(x), \ldots, g_n(x)\}, g_{n+1}(x)\}.$$

By assumption, the function

$$\bar{g}(x) := \max\{g_1(x), \ldots, g_n(x)\}$$

is continuous at $\bar{x}$. Thus

$$g(x) = \max\{\bar{g}(x), g_{n+1}(x)\}$$

is continuous at $\bar{x}$ as well, being the maximum of two functions that are both continuous at $\bar{x}$.

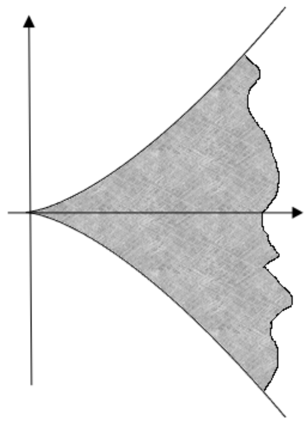
2. Consider the following problem:

$$\inf (x_1 + 1)^2 + x_2^2 \text{ subject to } -x_1^3 + x_2^2.$$

(a) Sketch the feasible region and hence solve the problem.
(b) Find multipliers λ_0 and λ satisfying the Fritz John conditions (2.3.6).
(c) Probe there exists no Lagrange multiplier vector for the optimal solution. Explain why not.

Date: October 14, 2015.

(a) The feasible region roughly looks like



The (obvious) solution of the problem is $x = (0, 0)$.

(b) The Fritz John condition requires us to find $\lambda_0 \geq 0$ and $\lambda_1 \geq 0$, not both simultaneously equal to 0, such that $\lambda_0 \nabla f(0) + \lambda_1 g(0) = 0$ with $f(x) = (x_1 + 1)^2 + x_2^2$ and $g(x) = -x_1^3 + x_2^2$. Thus $\nabla f(0) = (2, 0)$ and $\nabla g(0) = 0$, and we can choose $\lambda_0 = 0$ and $\lambda_1 > 0$ arbitrary.

(c) A Lagrange multiplier λ would satisfy the equation $(2, 0) + \lambda(0, 0) = 0$, which is obviously impossible. In this setting, the Mangasarian–Fromovitz constraint qualification does not hold because $\nabla g(0) = 0$, but at the same time 0 is no unconstrained minimizer of f.

5. (Cauchy–Schwarz and steepest descent) For a nonzero vector y in $\mathbb{E}$, use the Karush–Kuhn–Tucker conditions to solve the problem
$$\inf\{\langle y, x\rangle : \|x\|^2 \leq 1\}.$$
Deduce the Cauchy–Schwarz inequality.

We consider only the non-trivial situation $y \neq 0$.

We write $f(x) = \langle y, x\rangle$ and $g(x) = \|x\|^2 - 1$. Then $\nabla f(x) = y$ and $\nabla g(x) = 2x$. Thus the Karush–Kuhn–Tucker conditions for this problem require us to find $\lambda \geq 0$ and $x \in \mathbb{E}$ with $\|x\| \leq 1$ such that
$$y + 2\lambda x = 0.$$
Moreover, we have to choose $\lambda = 0$ if $\|x\|^2 < 1$. The only possible solution to this problem is $x = -y/\|y\|$ and $\lambda = \|y\|/2$.

As a consequence, we obtain that
$$\langle y, x\rangle \geq \langle y, -y/\|y\|\rangle = -\|y\|$$
for all x with $\|x\| \leq 1$, or, replacing x by $-x$, that
$$\langle y, x\rangle \leq \|y\|$$
for all x with $\|x\| \leq 1$. Consequently, we see that we have for every $x \neq 0$ the inequality
$$\langle y, x\rangle = \|x\|\langle y, x/\|x\|\rangle \leq \|x\|\|y\|,$$
which is precisely the Cauchy–Schwarz inequality. (Additionally, we obtain that equality holds if and only if $x = y$.)

7. Consider a matrix A in $\mathbb{S}^n_{++}$ and a real $b > 0$.
   (a) Assuming the problem
   $$\inf\{-\log\det X : \operatorname{tr} AX \le b,\ X \in \mathbb{S}^n_{++}\}$$
   has a solution, find it.
   (b) Repeat using the objective function $\operatorname{tr} X^{-1}$.
   (c) Prove the problems in parts (a) and (b) have optimal solutions. (Hint: Section 1.2, Exercise 14.)

(a) We write $f(X) = -\log\det X$ and $g(X) = \operatorname{tr} AX - b$ and try to solve the Karush–Kuhn–Tucker condition $\nabla f(X) + \lambda\nabla g(X) = 0$ with $\lambda \ge 0$ ($\lambda = 0$ if $g(X) < 0$) and $\operatorname{tr} AX \le b$. The main difficulty is the computation of ∇f and ∇g.

First we recall that the gradient of a differentiable function $h\colon \mathbb{S}^n_{++} \to \mathbb{R}$ at $X \in \mathbb{S}^n_{++}$ is defined to be the symmetric matrix $\nabla h(X)$ satisfying the equation
$$h'(X;Y) = \langle \nabla h(X), Y\rangle$$
for all $Y \in \mathbb{S}^n$, and the inner product on $\mathbb{S}^n$ is given by
$$\langle X, Y\rangle = \operatorname{tr} XY.$$
Since we can write g as $g(X) = \langle A, X\rangle - b$, it follows that $\nabla g(X) = A$.

The mapping f requires slightly more work, though. We first write f as $f(X) = -\log h(X)$ with $h(X) = \det X$. Then the chain rule implies that
$$\nabla f(X) = -\frac{1}{\det X}\nabla h(X).$$
For the calculation of ∇h, we first try to find $\nabla h(\mathrm{Id})$ (with $\mathrm{Id} \in \mathbb{R}^{n\times n}$ denoting the identity matrix).

Denote now by $E_{ij} \in \mathbb{R}^{n\times n}$ the matrix defined by $(E_{ij})_{kl} = \delta_{ik}\delta_{jl}$. Then a short computation shows that
$$h'(\mathrm{Id}; E_{ij}) = \lim_{t\to 0}\frac{1}{t}\big(\det(\mathrm{Id} + tE_{ij}) - 1\big) = \begin{cases} 0 & \text{if } i \ne j,\\ 1 & \text{if } i = j.\end{cases}$$
Writing $Y \in \mathbb{R}^{n\times n}$ as $\sum_{i,j} Y_{ij}E_{ij}$, we therefore conclude that
$$h'(\mathrm{Id};Y) = \sum_{i,j} Y_{ij} h'(\mathrm{Id};Y) = \sum_i Y_{ii} = \operatorname{tr} Y.$$
Now let $X \in \mathbb{S}^n$ be an arbitrary matrix with $\det(X) \ne 0$ and let $Y \in \mathbb{R}^{n\times n}$. Then
$$\begin{aligned} h'(X;Y) &= \lim_{t\to 0}\frac{1}{t}\big(\det(X + tY) - \det(X)\big)\\ &= \lim_{t\to 0}\frac{1}{t}\det(X)\big(\det(\mathrm{Id} + tX^{-1}Y) - 1\big)\\ &= \det(X)h'(\mathrm{Id}; X^{-1}Y)\\ &= \det(X)\operatorname{tr}(X^{-1}Y).\end{aligned}$$
Thus
$$h'(X) = \det(X)\operatorname{tr}(X^{-1}Y) = \det(X)\langle X^{-1}, Y\rangle,$$
implying that
$$\nabla h(X) = \det(X)X^{-1}.$$
As a consequence,
$$\nabla f(X) = -\frac{1}{\det X}\nabla h(X) = -X^{-1}.$$

In other words, the KKT-conditions require us to find $\lambda \geq 0$ and $X \in \mathbb{S}^n_{++}$ satisfying $\operatorname{tr} AX \leq b$ such that

$$-X^{-1} + \lambda A = 0. \tag{1}$$

Moreover, $\lambda = 0$ if $\operatorname{tr} AX < b$, which immediately implies that $\operatorname{tr} AX = b$. Equation (1) implies that $X = \sigma A^{-1}$ for some $\sigma > 0$. Now the condition

$$b = \operatorname{tr} AX = \operatorname{tr}(\sigma A A^{-1}) = n\sigma$$

implies that $\sigma = b/n$ or

$$X = \frac{b}{n} A^{-1}.$$

(b) This is somehow similar as (a), but now we have to compute $\nabla f(X)$ for $f(X) = \operatorname{tr} X^{-1}$. We first compute, for fixed $X \in \mathbb{S}^n_{++}$ and $Y \in \mathbb{S}^n$,

$$f'(X;Y) = \lim_{t \to 0} \frac{1}{t} \big(\operatorname{tr}(X+tY)^{-1} - \operatorname{tr}(X^{-1}) \big).$$

To that end we use the fact that, for t sufficiently small, we can write (Neumann series!)

$$\begin{aligned}(X+tY)^{-1} &= X^{-1}(\operatorname{Id} + tYX^{-1})^{-1} \\ &= X^{-1} \sum_{k=0}^{\infty} (-1)^k t^k (YX^{-1})^k = X^{-1} - tX^{-1}YX^{-1} + o(t).\end{aligned}$$

Thus

$$\operatorname{tr}(X+tY)^{-1} - \operatorname{tr}(X^{-1}) = -\operatorname{tr}(X^{-1}YX^{-1}) + o(t) = -\operatorname{tr}(X^{-2}Y) + o(t)$$

showing that

$$f'(X;Y) = -\operatorname{tr}(X^{-2}Y) = -\langle X^{-2}, Y \rangle$$

and thus

$$\nabla f(X) = -X^{-2}.$$

In this case, the KKT conditions thus require us to find $\lambda \geq 0$ and $X \in \mathbb{S}^n_{++}$ satisfying $\operatorname{tr} AX \leq b$ such that

$$-X^{-2} + \lambda A = 0, \tag{2}$$

and, again, $\lambda = 0$ if $\operatorname{tr} AX < b$. Now we obtain from (2) that

$$X = \sigma A^{-1/2}$$

with $A^{-1/2}$ being the inverse of the unique symmetric and positive definite root of A and $\sigma > 0$ chosen such that $\operatorname{tr}(AX) = b$. That is,

$$\sigma = \frac{b}{\operatorname{tr} A^{1/2}}$$

and

$$X = \frac{b}{\operatorname{tr} A^{1/2}} A^{-1/2}.$$

(c) We first note that a constrained optimization problem $\inf f(x)$ such that $g(x) \leq 0$ with f, g continuous has a solution if all the level sets of $f(x)+g(x)$ are compact. Indeed, in such a case we can choose a minimizing sequence x_k, i.e., a sequence x_k with $g(x_k) \leq 0$ and $f(x_k) \to \inf_{g(x) \leq 0} f(x)$. Then $f(x_k) + g(x_k) \leq \max_k f(x_k) < \infty$, and thus the compactness of the level sets of $f+g$ implies that there exists a subsequence $x_{k'}$ converging to some x. Now the continuity of f and g implies that $g(x) \leq \lim_{k'} g(x_{k'}) \leq 0$ and $f(x) \leq \lim_{k'} f(x_{k'}) = \inf_{g(x) \leq 0} f(x)$, showing that x is a solution of the constraint optimization problem.

- We start with the optimization problem

$$\inf\{-\log\det X : \operatorname{tr} AX \le b,\ X \in \mathbb{S}^n_{++}\}.$$

Here $f(X) = -\log\det X$ and $g(X) = \operatorname{tr} AX - b = \langle A, X\rangle - b$. Ignoring the constant b (which only changes the level of the level sets), we thus have to show that the level sets of the function

$$X \mapsto \langle A, X\rangle - \log\det X$$

are compact subsets of $\mathbb{S}^n_{++}$.

To that end, we note first that both f and g are continuous functions on $\mathbb{S}^n_{++}$ which implies that the level sets of $f+g$ are closed in $\mathbb{S}^n_{++}$. In order to show compactness, it remains to show that the level sets are bounded and that they stay away from the boundary of $\mathbb{S}^n_{+}$. In other words, given $c \in \mathbb{R}$, we have to show that

$$\inf\{\lambda_n(X) : f(X) + g(X) \le c\} > 0,$$
$$\sup\{\lambda_1(X) : f(x) + g(X) \le c\} < \infty.$$

In order to show this estimate, we need a lower bound for $\langle A, X\rangle$ in terms of $\lambda(A)$ and $\lambda(X)$. To that end, we note that Fan's inequality applied to $-A$ and X implies that

$$\langle -A, X\rangle \le \lambda(-A)^T\lambda(X) = \sum_{k=1}^n \lambda_k(-A)\lambda_k(X) = -\sum_{k=1}^n \lambda_{n-k}(A)\lambda_k(X),$$

and thus

$$\langle A, X\rangle \ge \sum_{k=1}^n \lambda_{n-k}(A)\lambda_k(X).$$

Now assume that $c \in \mathbb{R}$ and $f(X) + g(X) \le c$. Then

$$c \ge \langle A, X\rangle - \log\det(X) \ge \sum_{k=1}^n \lambda_{n-k}(A)\lambda_k(X) - \sum_{k=1}^n \log\lambda_k(X).$$

Denote now $d := \min_k \lambda_k(A) > 0$. Then this shows that

$$\sum_{k=1}^n \big(d\lambda_k(X) - \log\lambda_k(X)\big) \le c.$$

Noting that the function $t \mapsto dt - \log t$ takes the minimal value $\min_{t>0}(dt - \log t) = 1 + \log d$, this implies that

$$d\lambda_k(X) - \log\lambda_k(X) \le c - (n-1)(1+\log d) \tag{3}$$

for all k. Since $dt - \log t \to +\infty$ if either $t \to 0$ or $t \to \infty$, it follows that

$$\lambda_{\min} := \inf\{t > 0 : dt - \log t \le c - (n-1)(1+\log d)\} > 0$$

and

$$\lambda_{\max} := \sup\{t > 0 : dt - \log t \le c - (n-1)(1+\log d)\} < \infty.$$

Now (3) implies that

$$\lambda_{\min} \le \lambda_k(X) \text{ and } \lambda_k(X) \le \lambda_{\max}$$

for all k whenever $f(X)+g(X) \le c$, which proves the required bounds and thus the compactness of the level sets of $f + g$.

- Now we consider the optimization problem
$$\inf\left\{\operatorname{tr} X^{-1} : \operatorname{tr} AX \le b,\ X \in \mathbb{S}^n_{++}\right\}.$$
Here $f(X) = \operatorname{tr} X^{-1}$ and $g(X) = \langle A, X\rangle - b$. Noting that
$$f(X) = \operatorname{tr} X^{-1} = \sum_k \lambda_k^{-1}(X),$$
the argumentation is now almost identical to the previous one.