ESI 6420 HOMEWORK 4 SOLUTIONS

MAKSIM LEVENTAL

Time spent: 30 hours but I don't care because this is the last homework assignment I'll ever do in my life!

Collaborators: Chris Gianelli

1.1 Let $q = b - Ax$ to simplify notation. Then by LP duality if $\mathcal{P} := \min_y \left\{ d^\intercal y \middle| Dy \succeq q, y \succeq 0 \right\}$ is infeasible for some $x \in X$ the dual $\mathcal{D} := \max_u \left\{ u^\intercal q \middle| u^\intercal D \preceq d^\intercal, u \succeq 0 \right\}$ is unbounded. By characterization of unbounded LPs $\mathcal{D}$ being unbounded is equivalent there existing u satisfying the constraints and for which $u^\intercal q > 0$ i.e. u is an extreme ray. Therefore x is such that $(v^s)^\intercal (b - Ax) > 0$ for some $s \in \{1, \ldots, S\}$. Conversely if $\mathcal{P}$ is feasible for some $x \in X$ then, by LP duality, $\mathcal{D}$ is feasible and their optima are equal. Furthermore the optimal u it's the case that $u^\intercal q \le 0$ and u is an extreme point and hence $(u^p)^\intercal (b - Ax) \le 0$ for some $p \in \{1, \ldots, P\}$. Therefore both $(u^p)^\intercal (b - Ax) \le 0$ and $(v^s)^\intercal (b - Ax) \le 0$ for all u^p, v^s should be the case.

1.2 Let z be such that $z \ge u^\intercal (b - Ax) = (u^p)^\intercal (b - Ax)$ since the optimum for the inner maximization is at an extreme point of U. Then reformulated problem is

$$\begin{aligned} &\min_{x,z} \{c^\intercal x + z\} \\ &\text{s.t.}\ (v^s)^\intercal (b - Ax) \le 0 \,\forall s \\ &\quad (u^s)^\intercal (b - Ax) \le z \,\forall p \\ &\quad x \in X \end{aligned}$$

or

$$\begin{aligned} &\min_{x,z} \{c^\intercal x + z\} \\ &\text{s.t.}\ (v^s)^\intercal b - ((v^s)^\intercal A)\, x \le 0 \,\forall s \\ &\quad (u^p)^\intercal b - ((u^p)^\intercal A)\, x - z \le 0 \,\forall p \\ &\quad x \in X \end{aligned}$$

which is clearly linear in x, z.

1.3 The problem is

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{y}} - (3,7) \cdot \mathbf{x} + (1,2) \cdot \mathbf{y} \\ &\text{s.t.} - \mathbf{1} \cdot \mathbf{x} + -\mathbf{1} \cdot \mathbf{y} \ge -4.5 \\ &\quad - x_1 + y_1 \ge -1.5 \\ &\quad - x_2 + y_2 \ge -.5 \\ &\quad \mathbf{x} \in \mathbb{Z}^+ \times \mathbb{Z}^+ \\ &\quad \mathbf{y} \in \mathbb{R}^+ \times \mathbb{R}^+ \end{aligned}$$

In this instance

$$A = \begin{pmatrix} -1 & -1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}, D = \begin{pmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, b = \begin{pmatrix} -4.5 \\ -1.5 \\ -.5 \end{pmatrix}, c = (-3, -7), d = (1, 2)$$

Date: December 1, 2015.

Then

$$
\begin{aligned}
U &= \left\{u \in \mathbb{R}^3 \middle| D^{\intercal} u \leq d, u \geq 0\right\} \\
&= \left\{u \in \mathbb{R}^3 \middle| \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \preceq (1,2)\,, u \geq 0\right\}
\end{aligned}
$$

the polyhedron appears in figure 0.1.

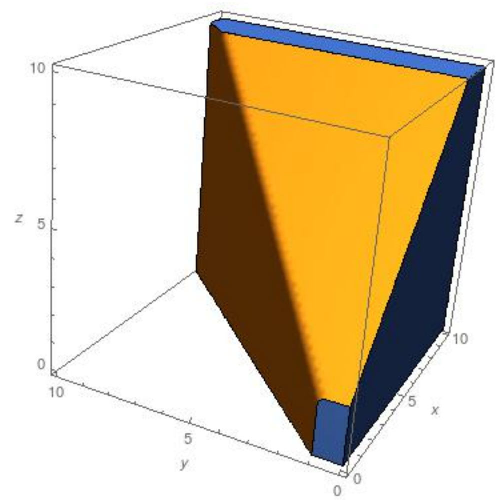

FIGURE 0.1. Bender Polyhedron

Clearly the vertices are

$$
\begin{aligned}
u^1 &= (0,0,0) \\
u^2 &= (0,1,0) \\
u^3 &= (0,0,2) \\
u^4 &= (0,1,2)
\end{aligned}
$$

and the extreme directions

$$
\begin{aligned}
v^1 &= (1,1,0) \\
v^2 &= (1,0,1) \\
v^3 &= (0,1,1) \\
v^4 &= (0,0,0)
\end{aligned}
$$

. Therefore the Bender reformulation is

$$
\begin{aligned}
&\min_{\mathbf{x},z} -(3,7)\cdot\mathbf{x}+z\\
&\text{s.t.}\ (0,0,0)\cdot\begin{pmatrix}-4.5\\-1.5\\-.5\end{pmatrix}-(0,0,0)\cdot\begin{pmatrix}-1&-1\\-1&0\\0&-1\end{pmatrix}\mathbf{x}\le 0\\
&\quad (0,1,0)\cdot\begin{pmatrix}-4.5\\-1.5\\-.5\end{pmatrix}-(0,1,0)\cdot\begin{pmatrix}-1&-1\\-1&0\\0&-1\end{pmatrix}\mathbf{x}\le 0\\
&\quad (0,0,2)\cdot\begin{pmatrix}-4.5\\-1.5\\-.5\end{pmatrix}-(0,0,2)\cdot\begin{pmatrix}-1&-1\\-1&0\\0&-1\end{pmatrix}\mathbf{x}\le 0\\
&\quad (0,1,2)\cdot\begin{pmatrix}-4.5\\-1.5\\-.5\end{pmatrix}-(0,1,2)\cdot\begin{pmatrix}-1&-1\\-1&0\\0&-1\end{pmatrix}\mathbf{x}\le 0\\
&\quad (1,1,0)\cdot\begin{pmatrix}-4.5\\-1.5\\-.5\end{pmatrix}-(1,1,0)\cdot\begin{pmatrix}-1&-1\\-1&0\\0&-1\end{pmatrix}\mathbf{x}-z\le 0\\
&\quad (1,0,1)\cdot\begin{pmatrix}-4.5\\-1.5\\-.5\end{pmatrix}-(1,0,1)\cdot\begin{pmatrix}-1&-1\\-1&0\\0&-1\end{pmatrix}\mathbf{x}-z\le 0\\
&\quad (0,1,1)\cdot\begin{pmatrix}-4.5\\-1.5\\-.5\end{pmatrix}-(0,1,1)\cdot\begin{pmatrix}-1&-1\\-1&0\\0&-1\end{pmatrix}\mathbf{x}-z\le 0\\
&\quad (1,0,0)\cdot\begin{pmatrix}-4.5\\-1.5\\-.5\end{pmatrix}-(1,0,0)\cdot\begin{pmatrix}-1&-1\\-1&0\\0&-1\end{pmatrix}\mathbf{x}-z\le 0\\
&\quad \mathbf{x}\in\mathbb{Z}^+\times\mathbb{Z}^+
\end{aligned}
$$

2.1 Claim: for a continuous $f:\mathbb{R}\to\mathbb{R}$ and $h>0$

$$f_h(x) := \frac{1}{2h}\int_{x-h}^{x+h} f(t)\,dt \ge f(x)$$

iff f is convex.

Proof. $\Leftarrow$ Suppose f is convex. Proceed by contradiction: suppose there exist h_0, x_0 such that $f(x_0) > f_{h_0}(x_0)$. Since f is convex there exists $g(x) = f(x_0) + m(x - x_0)$ such that $g \le f$. But then

$$f(x_0) = \frac{1}{2h}\int_{x_0-h_0}^{x_0+h_0} g(t)\,dt \le \frac{1}{2h}\int_{x_0-h_0}^{x_0+h_0} f(t)\,dt = f_{h_0}(x_0)$$

a contradiction.

$\Rightarrow$ Suppose $f_h(x) \ge f(x)$ for all h, x. Towards a contradiction suppose f is not convex. Then there exist λ_0, x_1, x_2 such that

$$f(\lambda_0 x_1 + (1-\lambda_0)x_2) > \lambda_0 f(x_1) + (1-\lambda_0) f(x_2)$$

where $\lambda_0 \in (0,1)$. Consider the function

$$F(\lambda) = f(\lambda x_1 + (1-\lambda)x_2) - (\lambda f(x_1) + (1-\lambda) f(x_2))$$

on $[0,1]$. Note that $F(0) = F(1) = 0$ and $F(\lambda_0) > 0$. Since F is continuous, being a linear function of f, it must achieve a maximum $F(\lambda^*)$ on $[0,1]$ (by the extreme value theorem) and $F(\lambda^*) > 0$ (since $F(\lambda_0) > 0$). Therefore there exists an h-ball around λ^* such that for

$\lambda \in [\lambda^* - h, \lambda^* + h]$ it's the case that $F(\lambda) > 0$ and without loss of generality [1]we can assume F is not constant on $[\lambda^* - h, \lambda^* + h]$. Then since $F(\lambda^*) \geq F(\lambda)$ for all $\lambda \in [\lambda^* - h, \lambda^* + h]$ and $F(\lambda^*) > F(\lambda)$ for at least one $\lambda \in [\lambda^* - h, \lambda^* + h]$ (otherwise F would be constant on $[\lambda^* - h, \lambda^* + h]$) we have that

$$2hF(\lambda^*) > \int_{\lambda^*-h}^{\lambda^*+h} F(\lambda)\, d\lambda$$

which is equivalent to

$$\begin{aligned}
&f(\lambda^* x_1 + (1-\lambda^*) x_2) - \\
&(\lambda^* f(x_1) + (1-\lambda^*) f(x_2)) > \frac{1}{2h} \int_{\lambda^*-h}^{\lambda^*+h} [f(\lambda x_1 + (1-\lambda) x_2) - (\lambda f(x_1) + (1-\lambda) f(x_2))]\, d\lambda \\
&\qquad = \frac{1}{2h'} \int_{\lambda^*-h}^{\lambda^*+h} f(\lambda x_1 + (1-\lambda) x_2)\, d\lambda - (\lambda^* f(x_1) + (1-\lambda^*) f(x_2))
\end{aligned}$$

Cancelling $-(\lambda^* f(x_1) + (1-\lambda^*) f(x_2))$ from both sides of the inequality we get that

$$f(\lambda^* x_1 + (1-\lambda^*) x_2) > \frac{1}{2h} \int_{\lambda^*-h}^{\lambda^*+h} f(\lambda x_1 + (1-\lambda) x_2)\, d\lambda = f_{h'}(\lambda^* x_1 + (1-\lambda^*) x_2)$$

contradicting that $f(x) \leq f_h(x)$ for all. Hence f must be convex. $\square$

2.2 Let $f(X) = -\log(\det(X))$.

(a) Claim: For $X, D \succeq 0$ and $X \succ 0$ and $g(t) = f(X + tD)$ it's the case that

$$g(t) = -\log\left(\det\left(\sqrt{X}\left(I + t\left(\sqrt{X}\right)^{-1} D \left(\sqrt{X}\right)^{-1}\right)\sqrt{X}\right)\right)$$

Proof. Firstly since $X \succ 0$ it's the case that X is full rank (all nonzero eigenvalues) and there exists a matrix $\sqrt{X}$ such that $\sqrt{X}\sqrt{X} = X$ and $\sqrt{X}$ is full rank[2]. Then $\left(\sqrt{X}\right)^{-1}$ exists and hence

$$\sqrt{X}\left((I+t)\left(\sqrt{X}\right)^{-1} D \left(\sqrt{X}\right)^{-1}\right)\sqrt{X} = X + tD$$

and so $g(t) = -\log(\det(X + tD)) = f(X + tD)$. $\square$

(b) Claim: $f(X)$ is convex.

Proof. Using the representation of g proven to be appropriate in part (a)

$$\begin{aligned}
g(t) &= -\log\left(\det\left(\sqrt{X}\right)\det\left(I + t\left(\sqrt{X}\right)^{-1} D \left(\sqrt{X}\right)^{-1}\right)\det\left(\sqrt{X}\right)\right) \\
&= -\log(\det(X)) - \log\left(\det\left(I + t\left(\sqrt{X}\right)^{-1} D \left(\sqrt{X}\right)^{-1}\right)\right)
\end{aligned}$$

[1] Why? If F is in fact constant on $[\lambda^* - h, \lambda^* + h]$ then we can take the minimum $h' > h$ such that either $\lambda^* - h' = 0$ or $\lambda^* + h' = 1$ and F cannot be constant on $[\lambda^* - h', \lambda^* + h']$. This is since, depending on whether h' is such that $\lambda^* - h' = 0$ or $\lambda^* + h' = 1$, either $F(\lambda^* - h') = F(0) = 0$ or $F(\lambda^* + h') = F(1) = 0$ (and F cannot equal zero on all $[\lambda^* - h, \lambda^* + h]$ since $F(\lambda_0) > 0$ and F is continuous).

[2]

Proof. $\sqrt{X} = Q\sqrt{\Sigma}Q^T$ where Q is the set of eigenvectors corresponding to X and $\sqrt{\Sigma} \succ 0$ since $\Sigma \succ 0$. $\square$

Let $Y = \left(\sqrt{X}\right)^{-1} D \left(\sqrt{X}\right)^{-1}$, which is PSD since D is PSD and $\left(\sqrt{X}\right)^{-1}$ is PD, and then

$$\begin{aligned} g(t) &= -\log(\det(X)) - \log(\det(I + tY)) \\ &= -\log(\det(X)) - \log\left(\prod_{i=1}^{n}(1 + t\lambda_i)\right) \\ &= -\log(\det(X)) - \sum_{i=1}^{n}\log(1 + t\lambda_i) \end{aligned}$$

So g is convex in t since it's the sum of a constant and convex functions of linear transformations of t. Hence $f(X)$ is convex since it is convex on every line. □

(1) Claim: Let $c \sim \mathcal{N}(\mu, \Sigma)$. Then assuming there exists x such that $P(c^\intercal x \geq \alpha) \geq \frac{1}{2}$

$$\begin{aligned} \max_{x\in\mathbb{R}^n} & P(c^\intercal x \geq \alpha) \\ \text{s.t.} & Fx \leq g \\ & Ax = b \end{aligned}$$

can be reformulated as a quadratic convex optimization problem.

Proof. Firstly since $c \sim \mathcal{N}(\mu, \Sigma)$ it's the case that $X = c^\intercal x \sim \mathcal{N}(\mu \cdot x, x^\intercal \Sigma x)$ and hence

$$P(X \geq \alpha) = P\left(\frac{X - \mu \cdot x}{\sqrt{x^\intercal \Sigma x}} \geq \frac{\alpha - \mu \cdot x}{\sqrt{x^\intercal \Sigma x}}\right) = P\left(Z \geq \frac{\alpha - \mu \cdot x}{\sqrt{x^\intercal \Sigma x}}\right)$$

where $Z \sim \mathcal{N}(0, 1)$. So the maximization problem is

$$\max_x \left[P\left(Z \geq \frac{\alpha - \mu \cdot x}{\sqrt{x^\intercal \Sigma x}}\right)\right]$$

Clearly maximizing this objective is equivalent to minimizing $\frac{\alpha - \mu \cdot x}{\sqrt{x^\intercal \Sigma x}}$. So the problem now is

$$\begin{aligned} \min_{x\in\mathbb{R}^n} & \frac{\alpha - \mu \cdot x}{\sqrt{x^\intercal \Sigma x}} \\ \text{s.t.} & Fx \preceq g \\ & Ax = b \end{aligned}$$

Alternatively we can maximize the reciprocal of the objective and hence solve the problem

$$\begin{aligned} \max_{x\in\mathbb{R}^n} & \frac{\sqrt{x^\intercal \Sigma x}}{\alpha - \mu \cdot x} \\ \text{s.t.} & Fx \preceq g \\ & Ax = b \end{aligned}$$

Alternatively (flipping the sign)

$$\begin{aligned} \min_{x\in\mathbb{R}^n} & \frac{\sqrt{x^\intercal \Sigma x}}{\mu \cdot x - \alpha} \\ \text{s.t.} & Fx \preceq g \\ & Ax = b \end{aligned}$$

The fact that there exists x_0 such that

$$P(c^\intercal x_0 \geq \alpha) \geq \frac{1}{2}$$

or

$$P\left(Z \geq \frac{\alpha - \mu \cdot x_0}{\sqrt{x_0^\intercal \Sigma x_0}}\right) \geq \frac{1}{2}$$

or

$$\frac{\alpha - \mu \cdot x_0}{\sqrt{x_0^\intercal \Sigma x_0}} \leq 0$$

or

$$\alpha - \mu \cdot x_0 \leq 0$$

implies

$$\left\{x \middle| Fx \preceq g, Ax = b, \mu \cdot x - \alpha \geq 0\right\} \neq \emptyset$$

Hence let $t = \frac{1}{\alpha - \mu \cdot x}$ and $y = xt$. Then an equivalent problem is

$$\begin{aligned} &\min_{y \in \mathbb{R}^n} \sqrt{y^\intercal \Sigma y} \\ &\text{s.t.} Fy \preceq gt \\ &\quad Ay = bt \\ &\quad \mu \cdot y - \alpha t = 1 \\ &\quad t \geq 0 \end{aligned}$$

Squaring the objective we get a convex program with a quadratic constraint. □

4.1 Let $x_1 = \ln(r)$ and $x_2 = \ln(h)$. Note this transformation is a bijection since $r, h > 0$ and $r \in (0, \infty)$ implies $x_1 \in (-\infty, \infty)$ and similarly for x_2. Further

$$2\pi\left(r^2 + rh\right) = 2\pi\left(e^{2x_1} + e^{x_1 + x_2}\right)$$

$$\pi r^2 h \geq V \iff 2x_1 + x_2 \geq \ln\left(\frac{V}{\pi}\right)$$

Hence the two problems are equivalent.

4.2 The new problem is a convex optimization problem because the objective is convex (being the sum of two convex functions of linear transformations) and the constraints are linear.

4.3 Since everything is differentiable we can just use calculus: let $f(x_1, x_2) = 2\pi\left(e^{2x_1} + e^{x_1 + x_2}\right)$. Then

$$\nabla f = \left(2e^{2x_1} + e^{x_1 + x_2}, e^{x_1 + x_2}\right)$$

Since this is never zero the constraint must be active (or the problem unbounded). On the constraint boundary

$$g(x_1) = f\left(\ln\left(\frac{V}{\pi}\right) - 2x_1, x_1\right) = 2\pi\left(e^{2x_1} + e^{x_1 + \ln\left(\frac{V}{\pi}\right) - 2x_1}\right) = 2\pi\left(e^{2x_1} + e^{\ln\left(\frac{V}{\pi}\right) - x_1}\right) = 2\pi\left(e^{2x_1} + \frac{V}{\pi}e^{-x_1}\right)$$

Then

$$g'(x_1) = 2\pi\left(2e^{2x_1} - \frac{V}{\pi}e^{-x_1}\right)$$

which is zero at

$$e^{2x_1} = \frac{V}{2\pi}e^{-x_1} \Rightarrow 2x_1 = \ln\left(\frac{V}{2\pi}\right) - x_1 \Rightarrow x_1 = \frac{1}{3}\ln\left(\frac{V}{2\pi}\right)$$

Therefore

$$x_2 = \ln\left(\frac{V}{\pi}\right) - \frac{2}{3}\ln\left(\frac{V}{2\pi}\right) = \ln(2) + \ln\left(\frac{V}{2\pi}\right) - \frac{2}{3}\ln\left(\frac{V}{2\pi}\right) = \ln(2) + \frac{1}{3}\ln\left(\frac{V}{2\pi}\right)$$

Hence

$$r = e^{\frac{1}{3}\ln\left(\frac{V}{2\pi}\right)} = \left(\frac{V}{2\pi}\right)^{\frac{1}{3}}$$

$$h = e^{\ln(2) + \frac{2}{3}\ln\left(\frac{V}{2\pi}\right)} = 2\left(\frac{V}{2\pi}\right)^{\frac{1}{3}}$$

5.1 Claim: There exists an extreme point $(\bar{x}, \bar{y})$ that solves the bilinear program.

Proof. **Full disclosure:** For this part I looked at Thieu's paper here http://journals.math.ac.vn/acta/pdf/198002106.pdf.

The problem can be restated as

$$\min_{x\in X}\left(c^{\intercal}x+\min_{y\in Y}\left(\left(d^{\intercal}+x^{\intercal}H\right)y\right)\right)$$

Note that $\left(\left(d^{\intercal}+x^{\intercal}H\right)y\right)$ is linear in y over the polyhedron Y hence the optimium is attained at an extreme point. Let $V\left(Y\right)$ be the set of extreme points Y. Then the problem can be restated again as

$$\min_{x\in X}\left(c^{\intercal}x+\min_{y\in V(Y)}\left(\left(d^{\intercal}+x^{\intercal}H\right)y\right)\right)$$

For each $\bar{y}\in V\left(Y\right)$ the function

$$g_{\bar{y}}\left(x\right)=\left(\bar{y}H+c^{\intercal}\right)x+d^{\intercal}\bar{y}$$

is a linear function of x. Hence the problem

$$\min_{x\in X}g\left(x\right)$$

is the minimization of a piecewise linear function of x over a polyhedron and therefore attains its minimum at an extreme point $\bar{x}$. Therefore $\left(\bar{x},\bar{y}\right)$ is a solution of the original bilinear problem and both $\bar{x},\bar{y}$ are extreme points of X,Y respectively. □

5.2 Claim: $\left(\hat{x},\hat{y}\right)$ is a local minimum iff $\left(c^{\intercal}+\hat{y}^{\intercal}H^{\intercal}\right)\left(x-\hat{x}\right)\geq0$ and $\left(d^{\intercal}+\hat{x}^{\intercal}H\right)\left(y-\hat{y}\right)\geq0$ for $x,y\in X,Y$ and $\left(c^{\intercal}+\hat{y}^{\intercal}H^{\intercal}\right)\left(x-\hat{x}\right)+\left(d^{\intercal}+\hat{x}^{\intercal}H\right)\left(y-\hat{y}\right)>0$ when $\left(x-\hat{x}\right)^{\intercal}H\left(y-\hat{y}\right)<0$.

Proof. $\Rightarrow$ Assume $\left(\hat{x},\hat{y}\right)$ is a local minimum. Then there exists some $\mathcal{N}_{\varepsilon}\left(\left(\hat{x},\hat{y}\right)\right)$ such that for all $\left(x,y\right)\in\mathcal{N}_{\varepsilon}\left(\left(\hat{x},\hat{y}\right)\right)$ it's the case that

$$c^{\intercal}x+d^{\intercal}y+x^{\intercal}Hy\geq c^{\intercal}\hat{x}+d^{\intercal}\hat{y}+\hat{x}^{\intercal}H\hat{y}$$

$\hat{x}$ being optimal implies that for any $\left(x,\hat{y}\right)\in\mathcal{N}_{\varepsilon}\left(\left(\hat{x},\hat{y}\right)\right)$

$$\begin{aligned}c^{\intercal}x+d^{\intercal}\hat{y}+x^{\intercal}H\hat{y}\geq c^{\intercal}\hat{x}+d^{\intercal}\hat{y}+\hat{x}^{\intercal}H\hat{y}\\ \Longleftrightarrow\\ c^{\intercal}\left(x-\hat{x}\right)+x^{\intercal}H\hat{y}\geq\hat{x}^{\intercal}H\hat{y}\\ \Longleftrightarrow\\ c^{\intercal}\left(x-\hat{x}\right)+\left(x-\hat{x}\right)^{\intercal}H\hat{y}\geq0\end{aligned}$$

Since $u^{\intercal}Av=\left(u^{\intercal}Av\right)^{\intercal}=v^{\intercal}A^{\intercal}u$ for all u,v,A

$$\begin{aligned}c^{\intercal}\left(x-\hat{x}\right)+\left(x-\hat{x}\right)^{\intercal}H\hat{y}\geq0\\ \Longleftrightarrow\\ c^{\intercal}\left(x-\hat{x}\right)+\hat{y}^{\intercal}H^{\intercal}\left(x-\hat{x}\right)\geq0\\ \Longleftrightarrow\\ \left(c^{\intercal}+\hat{y}^{\intercal}H^{\intercal}\right)\left(x-\hat{x}\right)\geq0\end{aligned}$$

Similarly for $\left(\hat{x},y\right)\in\mathcal{N}_{\varepsilon}\left(\left(\hat{x},\hat{y}\right)\right)$

$$\begin{aligned}d^{\intercal}y+\hat{x}^{\intercal}Hy\geq d^{\intercal}\hat{y}+\hat{x}^{\intercal}H\hat{y}\\ \Longleftrightarrow\\ d^{\intercal}\left(y-\hat{y}\right)+\hat{x}^{\intercal}Hy-\hat{x}^{\intercal}H\hat{y}\geq0\\ \Longleftrightarrow\\ d^{\intercal}\left(y-\hat{y}\right)+\hat{x}^{\intercal}H\left(y-\hat{y}\right)\geq0\\ \Longleftrightarrow\\ \left(d^{\intercal}+\hat{x}^{\intercal}H\right)\left(y-\hat{y}\right)\geq0\end{aligned}$$

Further if $(\hat{x}, \hat{y})$ is a local minimum

$$c^\intercal x + d^\intercal y + x^\intercal H y \geq c^\intercal \hat{x} + d^\intercal \hat{y} + \hat{x}^\intercal H \hat{y}$$
$$\iff$$
$$c^\intercal (x - \hat{x}) + d^\intercal (y - \hat{y}) + x^\intercal H y - \hat{x}^\intercal H \hat{y} \geq 0$$
$$\iff$$
$$c^\intercal (x - \hat{x}) + d^\intercal (y - \hat{y}) + (x^\intercal - \hat{x}^\intercal) H (y - \hat{y}) + (-2\hat{x}^\intercal H \hat{y} + x^\intercal H \hat{y} + \hat{x}^\intercal H y) \geq 0$$

Now since $(x - \hat{x})^\intercal H (y - \hat{y}) < 0$

$$c^\intercal (x - \hat{x}) + d^\intercal (y - \hat{y}) + (-2\hat{x}^\intercal H \hat{y} + x^\intercal H \hat{y} + \hat{x}^\intercal H y) > 0$$
$$\iff$$
$$c^\intercal (x - \hat{x}) + d^\intercal (y - \hat{y}) + (x^\intercal - \hat{x}^\intercal) H \hat{y} + \hat{x}^\intercal H (y - \hat{y}) > 0$$
$$\iff$$
$$(c^\intercal + \hat{y}^\intercal H^\intercal)(x - \hat{x}) + (d^\intercal + \hat{x}^\intercal H)(y - \hat{y}) > 0$$

Finally since, by part (1), since a solution exists at an extreme point, these inequalities hold for all $(x, y) \in X \times Y$.

Proof. $\Leftarrow$ Assume the inequalities hold for $(x, y) \in X \times Y$ and $(x - \hat{x})^\intercal H (y - \hat{y}) \geq 0$. Then

$$(c^\intercal + \hat{y}^\intercal H^\intercal)(x - \hat{x}) + (d^\intercal + \hat{x}^\intercal H)(y - \hat{y}) + (x - \hat{x})^\intercal H (y - \hat{y}) \geq 0$$
$$\iff$$
$$c^\intercal (x - \hat{x}) + d^\intercal (y - \hat{y}) \cancel{+ \hat{y}^\intercal H^\intercal x} - \hat{y}^\intercal H^\intercal \hat{x} \cancel{+ \hat{x}^\intercal H y} \cancel{- \hat{x}^\intercal H \hat{y}} + x^\intercal H y \cancel{- \hat{x}^\intercal H y} \cancel{- x^\intercal H \hat{y}} \cancel{+ \hat{x}^\intercal H \hat{y}} \geq 0$$
$$\iff$$
$$c^\intercal (x - \hat{x}) + d^\intercal (y - \hat{y}) + x^\intercal H y \geq \hat{y}^\intercal H^\intercal \hat{x}$$
$$\iff$$
$$c^\intercal x + d^\intercal y + x^\intercal H y \geq c^\intercal \hat{x} + d^\intercal \hat{y} + \hat{y}^\intercal H^\intercal \hat{x}$$

If for (x, y) it's the case that $(x - \hat{x})^\intercal H (y - \hat{y}) < 0$ then

$$(c^\intercal + \hat{y}^\intercal H^\intercal)(x - \hat{x}) + (d^\intercal + \hat{x}^\intercal H)(y - \hat{y}) > 0$$
$$\iff$$
$$(c^\intercal + \hat{y}^\intercal H^\intercal)(x - \hat{x}) + (d^\intercal + \hat{x}^\intercal H)(y - \hat{y}) + (x^\intercal - \hat{x}^\intercal) H (y - \hat{y}) \geq 0$$
$$\iff$$
$$c^\intercal x + d^\intercal y + x^\intercal H y \geq c^\intercal \hat{x} + d^\intercal \hat{y} + \hat{x}^\intercal H \hat{y}$$

$\square$

$\square$

5.3 Claim: $(\hat{x}, \hat{y})$ is a KKT point iff $(c^\intercal + \hat{y}^\intercal H^\intercal)(x - \hat{x}) \geq 0$ and $(d^\intercal + \hat{x}^\intercal H)(y - \hat{y}) \geq 0$ for $x, y \in X, Y$

Proof. Suppose $(\hat{x}, \hat{y})$ is a KKT point. By Gordan's theorem there exists no feasible (x, y) such that

$$\left(\nabla f(x, y)|_{(x,y)=(\hat{x},\hat{y})} \right)^\intercal (x - \hat{x}, y - \hat{y}) < 0$$

where $f(x, y) = c^\intercal x + d^\intercal y + x^\intercal H y$. Therefore for feasible all (x, y)

$$\left(\nabla f(x, y)|_{(x,y)=(\hat{x},\hat{y})} \right)^\intercal (x - \hat{x}, y - \hat{y}) \geq 0$$

Finally note that

$$\left(\nabla f(x, y)|_{(x,y)=(\hat{x},\hat{y})} \right)^\intercal = (c^\intercal + \hat{y}^\intercal H^\intercal, d^\intercal + \hat{x}^\intercal H)^\intercal$$

Since Gordan's theorem is iff the proof works in both directions. $\square$

5.4 For $f(x_1,x_2,y_1,y_2) = x_2 + y_1 + x_2y_1 - x_1y_2 + x_2y_2$ we have that $c = (0,1), d = (1,0)$ and

$$H = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$$

The point $(x_1,x_2,y_1,y_2) = (0,0,0,0)$ is a KKT point since

$$(c^\intercal + \hat{y}^\intercal H^\intercal)(x - \hat{x}) = (0,1)^\intercal x \geq 0$$

for $(x_1,x_2) \in X$ (since $x_2 \geq 0$ for $x \in X$) and

$$(d^\intercal + \hat{x}^\intercal H)(y - \hat{y}) = (1,0)^\intercal y \geq 0$$

for $(y_1,y_2) \in Y$ (since $y_1 \geq 0$ for $y \in Y$). The point $(0,0,0,0)$ is not a local minimum though since for $(x_1,x_2) = (a,0)$ and $(y_1,y_2) = (0,b)$ we have that

$$(x - \hat{x})^\intercal H (y - \hat{y}) = x^\intercal H y = x_2y_1 - x_1y_2 + x_2y_2 = -ab < 0$$

and yet

$$x_2 + y_1 = 0 + 0 \not> 0$$

and hence by part (2) $(a,0,0,b)$ is not a local minimum. On the other hand for $(x_1,x_2,y_1,y_2) = (3,0,1,5)$

$$(0,1)^\intercal (3,0) = 0 \geq 0$$

and

$$(1,0)^\intercal (1,5) = 1 \geq 0$$

and

$$(x - \hat{x})^\intercal H (y - \hat{y}) = x^\intercal H y = -3 \times 5 < 0$$

and

$$(0,1)^\intercal (3,0) + (1,0)^\intercal (1,5) = 1 > 0$$

The global minimum is indeed at $(3,0,1,5)$ with $f(3,0,1,5) = -14$.

6.1 Let P be the program

$$\begin{aligned} &\min_x - \ln(a^\intercal x) - \ln(b^\intercal x) \\ &\text{s.t.} \mathbf{1}^\intercal x = 1, x \succeq 0 \end{aligned}$$

which is equivalent

$$\begin{aligned} &\min_x - \ln(a^\intercal x) - \ln(b^\intercal x) \\ &\text{s.t.} \mathbf{1}^\intercal x - 1 = 0, -x \preceq 0 \end{aligned}$$

Since the objective is differentiable and each of the constraints is differentiable the KKT conditions are

$$\begin{aligned} -\frac{a}{a^\intercal x} - \frac{b}{b^\intercal x} + u_0\mathbf{1} - (u_1,\ldots,u_n) &= 0 \\ u_0(\mathbf{1}^\intercal x - 1) &= 0 \\ -(u_1x_1,\ldots,u_nx_n) &= 0 \\ (u_1,\ldots,u_n) &\succeq 0 \\ u_0 &\in \mathbb{R} \end{aligned}$$

with $\{\mathbf{1}^\intercal x - 1, -u_1x_1,\ldots,-u_nx_n\}$ linearly independent.

6.2 Let $x = \left(\frac{1}{2},0,\ldots,0,\frac{1}{2}\right)$. Since the problem is convex, if a point satisfies the KKT conditions then it's a global minimum. To wit

$$-\frac{a}{a^\intercal x} - \frac{b}{b^\intercal x} + u_0\mathbf{1} - (u_1,\ldots,u_n) = -\frac{a}{\frac{1}{2}(a_1+a_n)} - \frac{b}{\frac{1}{2}\left(\frac{1}{a_1}+\frac{1}{a_n}\right)} + u_0\mathbf{1} - (u_1,\ldots,u_n) = 0$$

$$u_0\left(\mathbf{1}^\intercal\left(\frac{1}{2},0,\ldots,0,\frac{1}{2}\right) - 1\right) = u_0 \times 0 = 0$$

$$-\left(\frac{1}{2}u_1,0,\ldots,0,\frac{1}{2}u_n\right) = 0$$

Therefore $u_1 = u_n = 0$ and otherwise $u_i \neq 0$. Hence for $u_0 = 2$ and $u_i = 2\left(1 - \frac{a_i + \frac{a_1 a_n}{a_i}}{a_1 + a_n}\right)$ for $i = 2, \ldots, n-1$

$$
\begin{aligned}
-\frac{a}{\frac{1}{2}(a_1+a_n)} - \frac{b}{\frac{1}{2}\left(\frac{1}{a_1}+\frac{1}{a_n}\right)} + u_0\mathbf{1} - (0, u_2 \ldots, u_{n-1}, 0) &= (u_0, u_0 - u_2, \ldots, u_0 - u_{n-1}, u_0) - \frac{2}{(a_1+a_n)}(a + a_1 a_n b) \\
&= (u_0, u_0 - u_2, \ldots, u_0 - u_{n-1}, u_0) \\
&\quad - \frac{2}{(a_1+a_n)}(a_1 + a_1 a_n b_1, a_2 + a_1 a_n b_2, \ldots, a_n + a_1 a_n b_n) \\
&= (u_0, u_0 - u_2, \ldots, u_0 - u_{n-1}, u_0) \\
&\quad - \frac{2}{(a_1+a_n)}(a_1 + a_n, a_2 + a_1 a_n b_2, \ldots, a_n + a_1) \\
&= (u_0, u_0 - u_2, \ldots, u_0 - u_{n-1}, u_0) - 2\left(1, \frac{a_2 + \frac{a_1 a_n}{a_2}}{a_1 + a_n}, \ldots, 1\right) \\
&= \left(u_0 - 2, u_0 - 2\left(\frac{a_2 + \frac{a_1 a_n}{a_2}}{a_1 + a_n}\right) - u_2, \ldots, u_0 - 2\right) \\
&= \left(0, 2\left(1 - \frac{a_2 + \frac{a_1 a_n}{a_2}}{a_1 + a_n}\right) - u_2, \ldots, 0\right) \\
&= 0
\end{aligned}
$$

All that remains is to show that $u_i \geq 0$ (since u_0 is free). Note that

$$\left(a_i - \frac{(a_1+a_n)}{2}\right)^2 \leq \left(a_n - \frac{(a_1+a_n)}{2}\right)^2$$

since $a_1 \leq a_i \leq a_n$ (i.e. any a_i is closer to the "middle" [average] than a_n). And I claim I'm done. Why?

$$\left(a_i - \frac{(a_1+a_n)}{2}\right)^2 = a_i^2 - a_i(a_1+a_n) + \left(\frac{(a_1+a_n)}{2}\right)^2$$

and

$$\left(a_n - \frac{(a_1+a_n)}{2}\right)^2 = \left(\frac{(a_1+a_n)}{2}\right)^2 - a_1 a_n$$

and therefore

$$a_i^2 - a_i(a_1+a_n) \leq -a_1 a_n$$

or

$$a_i^2 + a_1 a_n \leq a_i(a_1+a_n)$$

or

$$\frac{a_i + \frac{a_1 a_n}{a_i}}{a_1 + a_n} \leq 1$$

Hence

$$u_i = 2\left(1 - \frac{a_i + \frac{a_1 a_n}{a_i}}{a_1 + a_n}\right) \geq 2(1-1) = 0$$

Finally $z^* = -\ln\left(\frac{a_1}{2} + \frac{a_n}{2}\right) - \ln\left(\frac{1}{2a_1} + \frac{1}{2a_n}\right) = -\ln\left(\frac{(a_1+a_n)^2}{4a_1 a_n}\right)$.

6.3 Claim: for Hermitian pd A

$$2\sqrt{u^\intercal A u}\sqrt{u^\intercal A^{-1} u} \leq \sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}$$

Proof. Let $A = V\Sigma V^\intercal$ then $V^\intercal \Sigma V = A$ and $V^\intercal \Sigma^{-1} V = A^{-1}$

$$\sqrt{u^\intercal A u}\sqrt{u^\intercal A^{-1} u} = \sqrt{u^\intercal V^\intercal \Sigma V u}\sqrt{u^\intercal V^\intercal \Sigma^{-1} V u}$$

Let $v = Vu$. Therefore

$$\sqrt{u^\intercal A u}\sqrt{u^\intercal A^{-1} u} = \sqrt{v^\intercal \Sigma v}\sqrt{v^\intercal \Sigma^{-1} v}$$

Taking log of $\sqrt{v^\intercal \Sigma v}\sqrt{v^\intercal \Sigma^{-1} v}$

$$\ln\left(v^\intercal \Sigma v\right) + \ln\left(v^\intercal \Sigma^{-1} v\right) = \ln\left(\sum_{i=1}^{n} \lambda_i v_i^2\right) + \ln\left(\sum_{i=1}^{n} \frac{1}{\lambda_i} v_i^2\right)$$

Since V is orthonormal $\|v\|_2 = 1$ and hence it's the case that

$$\mathbf{1}^\intercal\left(v_1^2, \ldots, v_n^2\right) = \sum_{i=1}^{n} v_i^2 = \|v\|_2^2 = 1 \geq 1$$

Therefore by part (2) with $a_i = \lambda_i$ (with $a_i > 0$ since A is pd)

$$-\ln\left(\frac{\left(\lambda_1 + \lambda_n\right)^2}{4\lambda_1\lambda_n}\right) \leq -\ln\left(v^\intercal \Sigma v\right) - \ln\left(v^\intercal \Sigma^{-1} v\right)$$

or

$$\ln\left(v^\intercal \Sigma v\right) + \ln\left(v^\intercal \Sigma^{-1} v\right) \leq \ln\left(\frac{\left(\lambda_1 + \lambda_n\right)^2}{4\lambda_1\lambda_n}\right)$$

Taking exponentials (since the exponential is positive increasing) and square roots on both sides

$$\begin{aligned}\sqrt{\left(v^\intercal \Sigma v\right)\left(v^\intercal \Sigma^{-1} v\right)} &\leq \sqrt{\frac{\left(\lambda_1 + \lambda_n\right)^2}{4\lambda_1\lambda_n}} \\ &= \frac{1}{2}\frac{\left(\lambda_1 + \lambda_n\right)}{\sqrt{\lambda_1\lambda_n}} = \frac{1}{2}\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}\right)\end{aligned}$$

$\square$