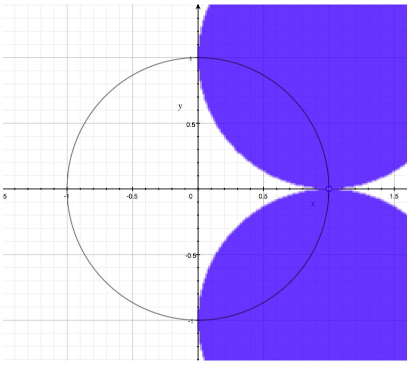
Exercise 5.26 [Boyd & Vandenberghe, 2004]

Ans: (a) The feasible set is the intersection of two disks, which contains only one point $x^* = (1, 0)$, and the optimum is $p^* = 1$.



(b) Indeed $x^* = (1, 0)$ satisfies the constraints. The gradient of Lagrangian is

$$\begin{aligned}
\nabla_x L(x^*, \lambda_1, \lambda_2) &= \nabla_x \left(\|x^*\|_2^2 + \lambda_1 \left(\left\| x^* - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\|^2 - 1 \right) + \lambda_2 \left(\left\| x^* - \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\|^2 - 1 \right) \right) \\
&= 2x^* + 2\lambda_1 \left(x^* - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) + 2\lambda_2 \left(x^* - \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right) \\
&= \begin{bmatrix} 2 \\ -2\lambda_1 + 2\lambda_2 \end{bmatrix}
\end{aligned}$$

Notice that there exists no λ_1^*, λ_2^* that $\nabla_x L(x^*, \lambda_1^*, \lambda_2^*) = 0$ since the first entry of it is constantly 2.

(c) From the above computation, we know the Lagrange dual function is not much but imposing $2x + 2\lambda_1 \left(x - \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right] \right) + 2\lambda_2 \left(x - \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right] \right) = 0$ in the Lagrangian; this gives us

$$\begin{aligned}
\bar{x} &= \frac{1}{1 + \lambda_1 + \lambda_2} \begin{bmatrix} \lambda_1 + \lambda_2 \\ \lambda_1 - \lambda_2 \end{bmatrix} \\
g(\lambda_1, \lambda_2) &= L(\bar{x}, \lambda_1, \lambda_2) \\
&= \frac{-(\lambda_1 - \lambda_2)^2 + \lambda_1 + \lambda_2}{1 + \lambda_1 + \lambda_2}
\end{aligned}$$

(may glory be to the great Wolfram Alpha) with the assumption that $1 + \lambda_1 + \lambda_2 > 0$ (we impose this to ensure Hessian of Lagrangian is positive definite to get infimum). The dual problem is

$$\begin{array}{ll}
\text{maximize} & \dfrac{-(\lambda_1 - \lambda_2)^2 + \lambda_1 + \lambda_2}{1 + \lambda_1 + \lambda_2} \\
\text{subject to} & \lambda_1, \lambda_2 \geq 0
\end{array}$$

with an implicit constraint $1 + \lambda_1 + \lambda_2 > 0$. Looking at the negative squared term we notice $\lambda_1 = \lambda_2$ needs to hold when maximized. The problem becomes maximizing $\frac{t}{1+t}$ where $t = \lambda_1 + \lambda_2 \geq 0$. This is bounded above by 1 yet can't be maximized. Strong duality does not hold for this problem. (Notice that we failed Slater's condition for this one.) $\square$

Exercise 5.29 [Boyd & Vandenberghe, 2004]

Ans: Let $A = \begin{bmatrix} -3 & & \\ & 1 & \\ & & 2 \end{bmatrix}$; then take the Lagrangian and its gradient w.r.t. x,

$$\begin{aligned} L(x,\nu) &= x^T A x + 2\mathbf{1}^T x + \nu(\|x\|^2 - 1) \\ \nabla_x L(x,\nu) &= 2Ax + 2\mathbf{1} + 2\nu x \end{aligned}$$

KKT condition suggests that $\nabla_x L(x^*,\nu^*) = 0$, *i.e.*

$$x = (A+\nu I)^{-1}\mathbf{1} = \begin{bmatrix} \frac{1}{\nu-3} \\ \frac{1}{\nu+1} \\ \frac{1}{\nu+2} \end{bmatrix}$$

and $\|x^*\|^2 = 1$, *i.e.*

$$\left(\frac{1}{\nu-3}\right)^2 + \left(\frac{1}{\nu+1}\right)^2 + \left(\frac{1}{\nu+2}\right)^2 = 1$$

Again, may glory be to the great Wolfram Alpha, this is equivalent to solving the sextic[1] polynomial equation $\nu^6 - 17\nu^4 - 12\nu^3 + 49\nu^2 + 48\nu - 13 = 0$ which has 4 real roots $\nu^* \approx -3.14929, 0.223509, 1.8919, 4.03523$. Plug them into our objective function

```
>> nu = [-3.14929, 0.223509, 1.8919, 4.03523];
>> x = @(nu) [1/(nu-3);1/(nu+1);1/(nu+2)];
>> A = diag([-3,1,2]);
>> obj = @(nu) x(nu)'*A*x(nu) + 2*sum(x(nu));
>> for  i=1:4
result(i) = obj(nu(i));
end
>> result

result =

   -1.3447    2.4972   -2.7910   -0.0444
```

The optimum is $p^* \approx -2.7910$ which corresponds to $\nu^* \approx 1.8919$. □

[1] OSX autocorrection does not recognize this word. It describes a polynomial to have degree 6.

Exercise 5.30 [Boyd & Vandenberghe, 2004]

Ans: Take the Lagrangian and its gradient w.r.t. X,

$$\begin{aligned} L(X,\nu) &= tr(X) - \log\det X + \nu^T(Xs-y) \\ \nabla_X L(X,\nu) &= diag(X) - \frac{1}{\det X} adj(X) + \nu s^T \\ &= diag(X) - X^{-1} + \nu s^T \end{aligned}$$

To examine if $X^* = I + yy^T - \frac{1}{\|s\|^2} ss^T$ is the optimal solution, first observe that $X^* s = s + y - s = y$, the equality constraint is satisfied. We use the Sherman-Morrison formula

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1 + v^TA^{-1}u}$$

to help compute $(X^*)^{-1}$:

$$\begin{aligned} (I + yy^T)^{-1} &= I - \frac{yy^T}{1 + \|y\|^2} \\ \left(I + yy^T - \frac{1}{\|s\|^2} ss^T\right)^{-1} &= (I + yy^T)^{-1} + \frac{(I + yy^T)^{-1}ss^T(I + yy^T)^{-1}}{\|s\|^2 - s^T(I + yy^T)^{-1}s} \\ &= I - \frac{yy^T}{1 + \|y\|^2} + \frac{(I - \frac{yy^T}{1+\|y\|^2})ss^T(I - \frac{yy^T}{1+\|y\|^2})}{\|s\|^2 - s^T(I - \frac{yy^T}{1+\|y\|^2})s} \\ &= I - \frac{yy^T}{1 + \|y\|^2} + \frac{(I - \frac{yy^T}{1+\|y\|^2})(ss^T - \frac{sy^T}{1+\|y\|^2})}{1/(1 + \|y\|^2)} \\ &= I - \frac{yy^T}{1 + \|y\|^2} + \frac{ss^T - \frac{ys^T}{1+\|y\|^2} - \frac{sy^T}{1+\|y\|^2} + \frac{yy^T}{(1+\|y\|^2)^2}}{1/(1 + \|y\|^2)} \\ &= I - \frac{yy^T}{1 + \|y\|^2} + (1 + \|y\|^2)ss^T - ys^T - sy^T + \frac{yy^T}{1 + \|y\|^2} \\ &= I + (1 + \|y\|^2)ss^T - ys^T - sy^T \\ &= I + ss^T - sy^T \end{aligned}$$

(I have a feeling that I took a huge detour... anyways I verified the formula and indeed $(I + ss^T - sy^T)X^* = I$) Hence the gradient of Lagrangian condition becomes

$$\begin{aligned} \nabla_X L(X^*,\nu) &= diag(X^*) - I - ss^T + sy^T + \nu s^T \\ &= diag(yy^T) - \frac{1}{\|s\|^2} diag(ss^T) - ss^T + sy^T + \nu s^T = 0 \end{aligned}$$

But inspecting the diagonal entries of this matrix equation, we found

$$y_i^2 - \frac{1}{\|s\|^2} s_i^2 - s_i^2 + s_i y_i + \nu_i s_i = 0$$

Temporarily let $\nu_i = \frac{1}{s_i}\left(-y_i^2 + \frac{1}{\|s\|^2} s_i^2 + s_i^2 - s_i y_i\right)$. By inspection we know this will satisfy that $\nabla_X L(X^*,\nu) = 0$, the KKT conditions are satisfied, and X^* is a optimal solution. $\square$

Additional Exercise 4.10 [Boyd & Vandenberghe, 2017]

Ans: (a) The Lagrange function

$$\begin{aligned} L(x,\nu) &= \|Ax-b\|_2^2 + x^T diag(\nu)x - \mathbf{1}^T\nu \\ &= b^Tb + x^T(A^TA + diag(\nu))x - 2b^TAx - \mathbf{1}^T\nu \end{aligned}$$

will be unbounded below if $A^TA + diag(\nu) \not\succeq 0$; hence under the assumption that $A^TA + diag(\nu) \succeq 0$, we can take derivative w.r.t. x in order to find $g(\nu)$.

$$\begin{aligned} \nabla_x L(x,\nu) &= 2(A^TA + diag(\nu))x - 2A^Tb = 0 \\ x &= (A^TA + diag(\nu))^{-1}A^Tb \\ g(\nu) &= b^Tb - \mathbf{1}^T\nu - b^TA(A^TA + diag(\nu))^{-1}A^Tb \end{aligned}$$

The dual problem is therefore a SDP:

$$\begin{array}{ll} \text{maximize} & b^Tb - \mathbf{1}^T\nu - b^TA(A^TA + diag(\nu))^{-1}A^Tb \\ \text{subject to} & A^TA + diag(\nu) \succeq 0 \end{array}$$

(b) Since there's a SDP constraint, the dual variable $Z \in \mathbf{S}^n$ and the Lagrangian is

$$\begin{aligned} L(\nu,Z) &= -b^Tb + \mathbf{1}^T\nu + b^TA(A^TA + diag(\nu))^{-1}A^Tb - Z : (A^TA + diag(\nu)) \\ \nabla_\nu L(\nu,Z) &= \mathbf{1} + (A^TA + diag(\nu))^{-1}A^Tb - diag(Z) \end{aligned}$$

This is from the fact that $\delta(A^{-1}) = A^{-1}\delta A A^{-1}$,

$$\frac{\partial(A^TA + diag(\nu))^{-1}}{\partial\nu_i} = (A^TA + diag(\nu))^{-1}E_{ii}(A^TA + diag(\nu))^{-1}$$

where E_{ii} denotes the matrix with 1 at its (i,i)-th entry and 0 for the rest. The gradient of Lagrangian can be zero if

$$\begin{aligned} A^Tb &= (A^TA + diag(\nu))(diag(Z) - \mathbf{1}) \\ \nu_i &= (A_{ji}b_j - A_{ji}A_{jk}Z_{kk} + A_{ji}A_{jk})/(Z_{ii} - 1) \end{aligned}$$

Of course index notation can't help us better the understanding...

Additional Exercise 4.14 [Boyd & Vandenberghe, 2017]

Ans: (a) We verify that for $x = (1/2, 0, \cdots, 0, 1/2)$, indeed $x \succeq 0, \mathbf{1}^T x = 1$; due to complementary slackness, we impose $\lambda_1 = \lambda_n = 0$. The gradient of Lagrangian is then

$$\begin{aligned}
\nabla_x L(x, \lambda, \nu) &= \nabla_x \left(-\log(a^T x) - \log(b^T x) - \lambda^T x + \nu(\mathbf{1}^T x - 1) \right) \\
&= -\frac{1}{a^T x} a - \frac{1}{b^T x} b - \lambda + \nu \mathbf{1} \\
&= \begin{bmatrix} \nu - \frac{2a_1}{a_1+a_n} - \frac{2b_1}{b_1+b_n} \\ -\lambda_2 + \nu - \frac{2a_2}{a_1+a_n} - \frac{2b_2}{b_1+b_n} \\ \vdots \\ -\lambda_{n-1} + \nu - \frac{2a_{n-1}}{a_1+a_n} - \frac{2b_{n-1}}{b_1+b_n} \\ \nu - \frac{2a_n}{a_1+a_n} - \frac{2b_n}{b_1+b_n} \end{bmatrix}
\end{aligned}$$

This can be solved since

$$\begin{aligned}
\frac{2a_1}{a_1 + a_n} + \frac{2b_1}{b_1 + b_n} &= \frac{2a_1}{a_1 + a_n} + \frac{2 \cdot \frac{1}{a_1}}{\frac{1}{a_1} + \frac{1}{a_n}} \\
&= \frac{2a_1}{a_1 + a_n} + \frac{2a_n}{a_1 + a_n} = 2 \\
&= \frac{2a_n}{a_1 + a_n} + \frac{2b_n}{b_1 + b_n}
\end{aligned}$$

we can let $\nu = 2$ then the first and n-th entries of the above gradient cancel. Let λ_i for $i = 2, \cdots, n-1$ accordingly to cancel the terms, then

$$\begin{aligned}
\lambda_i &= \nu - \frac{2a_i}{a_1 + a_n} - \frac{2b_i}{b_1 + b_n} \\
&= 2 - \frac{2a_i}{a_1 + a_n} - \frac{2b_i}{b_1 + b_n} \geq 0
\end{aligned}$$

(since a_i's and b_i's are in descending/ascending order) and the condition $\lambda \succeq 0$ is satisfied.

(b) Consider the eigendecomposition $A = Q \Lambda Q^T$ and let $y = Q^T u$, $x = y : y$ (*i.e.* $x_i = y_i^2$ for $i = 1, \cdots, n$), then the LHS of the inequality becomes

$$2 \left(u^T A u \right)^{1/2} \left(u^T A^{-1} u \right)^{1/2} = 2 \left(y^T \Lambda y \right)^{1/2} \left(y^T \Lambda^{-1} y \right)^{1/2} = 2(a^T x)^{1/2} (b^T x)^{1/2}$$

Due to part (a), we know the negative logarithm of this expression is bounded from below by substituting in $x^* = (1/2, 0, \cdots, 0, 1/2)$. Since negative logarithm is known to reverse order, we know now

$$
\begin{aligned}
2\left(u^T A u\right)^{1/2}\left(u^T A^{-1} u\right)^{1/2} &= 2(a^T x)^{1/2}(b^T x)^{1/2} \\
&\leq 2(a^T x^*)^{1/2}(b^T x^*)^{1/2} \\
&= 2\left(\frac{a_1 + a_n}{2}\right)^{1/2}\left(\frac{b_1 + b_n}{2}\right)^{1/2} \\
&= 2\left(\frac{\lambda_1 + \lambda_n}{2}\right)^{1/2}\left(\frac{\lambda_n + \lambda_n}{2\lambda_1\lambda_n}\right)^{1/2} \\
&= \frac{\lambda_1 + \lambda_n}{\lambda_1^{1/2}\lambda_n^{1/2}} \\
&= \sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}
\end{aligned}
$$

□

Additional Exercise 4.17 [Boyd & Vandenberghe, 2017]

Ans: (a)
"$tr(AX^*) \geq f(A)$": Take the unit eigenvectors $v_1, \cdots, v_r$ corresponding to $\lambda_1(A), \cdots, \lambda_r(A)$ and let $X = \sum_{i=1}^{r} v_i v_i^T$, then

$$tr(AX) = \sum_{i=1}^{r} tr(Av_i v_i^T) = \sum_{i=1}^{r} tr(v_i^T A v_i) = \sum_{i=1}^{r} \lambda_i(A)$$

Now enough to show that X satisfies the constraints of the problem. First we observe

$$tr(X) = \sum_{i=1}^{r} tr(v_i v_i^T) = \sum_{i=1}^{r} tr(v_i^T v_i) = \sum_{i=1}^{r} \|v_i\|_2^2 = r$$

Secondly, X is certainly semi-positive definite (it's the sum of r semi-positive definite matrices), and for any $x \in \mathbb{R}^n$,

$$x^T(I - X)x = \sum_{i=1}^{r} \|x\|^2 - (x^T v_i)^2 \geq 0$$

due to Cauchy's inequality, and $X \preceq I$.

"$tr(AX^*) \leq f(A)$": Suppose $X \in \mathbf{S}^n$ satisfies the constraints $tr(X) = r, 0 \preceq X \preceq I$; take the eigendecomposition $X = UMU^T$ with the diagonals of M being $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$, then from the constraints we know $\sum_{i=1}^{n} \mu_i = r$ and $0 \leq \mu_i \leq 1$ for all $i = 1, \cdots, n$. Now from the cyclic property of trace,

$$\begin{aligned} tr(AX) &= tr(AUMU^T) \\ &= tr(U^T AUM) \\ &= diag(U^T AU) : M \\ &= \sum_{i=1}^{n} \mu_i u_i^T A u_i \end{aligned}$$

(Here $A : B = \sum_{i,j=1}^{n} a_{ij} b_{ij}$ denotes the entry-wise product.) From the ordering that $1 \geq \mu_1 \geq \mu_2 \geq \cdots \geq \mu_n \geq 0$ we notices this is bounded by $\sum_{i=1}^{n} \mu_i \lambda_i(A)$ due to min-max principle and that u_i's form a basis. Along with the constraint that $\sum_{i=1}^{n} \mu_i = r$ we can see now

$$\text{tr}(AX) = \sum_{i=1}^{n} \mu_i u_i^T A u_i \leq \sum_{i=1}^{n} \mu_i \lambda_i(A) \leq \sum_{i=1}^{r} \lambda_i(A) = f(A)$$

(b) Utilize the result from part (a); denote the feasible set of problem in (a) by

$$S = \{X \in \mathbf{S}^n \mid tr(X) = r, 0 \preceq X \preceq I\}$$

. Take $\theta \in [0, 1]$ and $A, B \in \mathbf{S}^n$, then

$$\begin{aligned}
f(\theta A + (1-\theta)B) &= \max_{X \in S} tr((\theta A + (1-\theta)B)X) \\
&= \max_{X \in S} \theta tr(AX) + (1-\theta) tr(BX) \\
&\leq \theta \max_{X \in S} tr(AX) + (1-\theta) \max_{Y \in S} tr(BY) \\
&= \theta f(A) + (1-\theta) f(B)
\end{aligned}$$

(c) With the above derivations, we now know the problem is equivalent to

$$\begin{array}{ll}
\text{minimize} & t \\
\text{subject to} & y_0 + x^T y \leq t \\
& y_i = tr(A_i X), \text{ for } i = 0, 1, \cdots, m \\
& tr(X) = r \\
& 0 \preceq X \preceq I
\end{array}$$

Note here the variable $y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$ excludes the digit y_0. $\square$

Additional Exercise 4.18 [Boyd & Vandenberghe, 2017]

I accidentally did this problem...
Ans: (a) The maximum of several convex functions is convex, and taking linear combination of two convex functions with positive coefficients is still convex.

(b) Denote the Lagrange dual function for problem (16) by

$$g_{16}(\lambda) = \inf_x f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)$$

where $\lambda = [\lambda_1, \cdots, \lambda_m]^T$; then the Lagrangian and the Lagrange dual function for problem (19) is (with one more dual variable λ_0 due to the constraint $y \leq 0$)

$$\begin{aligned}
L_{19}(x, y, \lambda_0, \lambda) &= f_0(x) + ty + \sum_{i=1}^m \lambda_i (f_i(x) - y) + \lambda_0 y \\
&= f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \left(t + \lambda_0 - \sum_{i=1}^m \lambda_i\right) y \\
g_{19}(\lambda_0, \lambda) &= \inf_{x,y} L(x, y, \lambda) \\
&= g_{16}(\lambda) + \inf_y \left(t + \lambda_0 - \sum_{i=1}^m \lambda_i\right) y \\
&= \begin{cases} g_{16}(\lambda), & t + \lambda_0 - \sum\limits_{i=1}^m \lambda_i = 0 \\ -\infty, & \text{else} \end{cases}
\end{aligned}$$

(c) Assume $t > \mathbf{1}^T \lambda^*$, then $g_{19}(\mathbf{1}^T \lambda^* - t, \lambda)$ can attain its maximum $g_{16}(\lambda^*)$. Under the assumption of Slater's condition, optimal solution (x^*, y^*) to problem (19) (which is equivalent to problem (18) if we take only x^*) will give the same optimal value $g_{19}(\lambda^*)$ as problem (16). Since problem (18) is "weaker" than problem (16) (*i.e.* optimal solution to problem (16) will also solve problem (18)), we know then x^* solves problem (16). $\square$