HW-2

author: Ilya Drobyshevskiy

Convexity

1. Show, that $\mathbf{conv}\{xx^\top : x \in \mathbb{R}^n, \|x\| = 1\} = \{A \in \mathbb{S}^n_+ : \text{tr}(A) = 1\}$. **Solution**: Let's check if $\text{tr}(\theta_1 x_1 x_1^\top + \ldots + \theta_n x_n x_n^\top) = 1$, where $\sum_{i=1}^n \theta_i = 1$

$$\text{tr}(\theta_1 x_1 x_1^\top + \ldots + \theta_n x_n x_n^\top) = \theta_1 \text{tr}(x_1 x_1^\top) + \ldots + \theta_n \text{tr}(x_n x_n^\top) =$$

   Note, that

$$\text{tr}(xx^\top) = \text{tr}(x^\top x) = \|x\|_2^2 = 1$$

   So

$$= \theta_1 + \ldots + \theta_n = 1$$

   Also we have to check if

$$\theta_1 x_1 x_1^\top + \ldots + \theta_n x_n x_n^\top \geqslant 0$$

   If we show that $xx^\top \geqslant 0$ sum of positive semi-definite matrix

$$yxx^\top y = \|x^\top y\|_2^2 \geqslant 0$$

2. Prove that the set of $\{x \in \mathbb{R}^2 \mid e^{x_1} \leq x_2\}$ is convex. **Solution**: If a set is convex then $\forall x, y \in S, \theta \in [0, 1]$

$$z = \theta x + (1 - \theta) y \in S$$

   We have to show that $e^{z_1} \leqslant z_2$

$$e^{z_1} = e^{\theta x_1 + (1-\theta) y_1} \leqslant \theta e^{x_1} + (1 - \theta) e^{y_1} \leqslant \theta x_2 + (1 - \theta) y_2 = z_2$$

   Here we use that exp is a convex function.

3. Show that the set of directions of the non-strict local descending of the differentiable function in a point is a convex cone. **Solution**: The set we work with is

$$D = \{d \in \mathbb{R}^n | \nabla f(x)^\top d \leqslant 0\}$$

   For proof that it is a convex set we have to show that linear hull lies in D and D is a cone

   1. $\forall d_1, d_2 \in D, \theta_1, \theta_2 > 0$

$$\nabla f(x)^\top (\theta_1 d_1 + \theta_2 d_2) = \theta_1 \nabla f(x)^\top d_1 + \theta_2 \nabla f(x)^\top d_2 \leqslant 0$$

   2.$\forall d \in D, \theta \in [0, 1]$

$$\nabla f(x)^\top (\theta d) = \theta \nabla f(x)^\top d \leqslant 0$$

4. Is the following set convex

$$S = \left\{a \in \mathbb{R}^k \mid p(0) = 1, |p(t)| \leq 1 \text{ for } \alpha \leq t \leq \beta\right\},$$

where

$$p(t) = a_1 + a_2 t + \ldots + a_k t^{k-1} \ ?$$

Solution: Let's $x, y \in S, \theta \in [0, 1]$ and $z = \theta x + (1-\theta) y$.

$$p_x(t) = x_1 + x_2 t + \ldots + x_k t^{k-1}$$

$$p_y(t) = y_1 + y_2 t + \ldots + y_k t^{k-1}$$

$$\implies p_z(t) = \theta p_x(t) + (1-\theta) p_y(t)$$

Firstly let's check that $p_z(0) = 1$

$$p_z(0) = \theta p_x(0) + (1-\theta) p_y(0) = \theta + (1-\theta) = 1$$

Now lets check that $|p_z(t)| \leqslant 1$

$$|p_z(t)| = |\theta p_x(t) + (1-\theta) p_y(t)| \leqslant \theta |p_x(t)| + (1-\theta)|p_y(t)| \leqslant \theta + (1-\theta) = 1$$

5. Consider the function $f(x) = x^d$, where $x \in \mathbb{R}_+$. Fill the following table with ✓ or X. Explain your answers

d	Convex	Concave	Strictly Convex	μ-strongly convex
$-2, x \in \mathbb{R}_{++}$	✓	X	✓	X
$-1, x \in \mathbb{R}_{++}$	✓	X	✓	X
0	✓	✓	X	X
0.5	X	✓	X	X
1	✓	✓	X	X
$\in (1; 2)$	✓	X	✓	X
2	✓	X	✓	✓
> 2	✓	X	X	X

Solution:

- To check for convexity it was used that $f''(x) \geqslant 0$
- To check for concavity it was used that $f''(x) \leqslant 0$
- To check for strict convexity it was used that $f''(x) > 0$
- To check for strong convexity it was used that $f''(x) \geqslant \mu > 0$

6. Prove that the entropy function, defined as

$$f(x) = -\sum_{i=1}^{n} x_i \log(x_i),$$

with $\text{dom}(f) = \{x \in \mathbb{R}^n_{++} : \sum_{i=1}^n x_i = 1\}$, is strictly concave. **Solution**: Let's prove that $g(x) = -f(x)$ is strictly convex:

$$g(\theta x + (1-\theta)y) = \sum_{i=1}^n (\theta x_i + (1-\theta)y_i)\log(\theta x_i + (1-\theta)y_i) <$$

Now let's prove that $x\log x$ is strictly convex:

$$\frac{\partial}{\partial x^2}(x\log x) = \frac{\partial}{\partial x}(\log x + 1) = \frac{1}{x} > 0$$

Now that we can finish the proof

$$< \theta\sum_{i=1}^n x_i\log x_i + (1-\theta)\sum_{i=1}^n y_i\log y_i = \theta g(x) + (1-\theta)g(y)$$

So if g is strictly convex then f is strictly concave.

7. Show, that the function $f : \mathbb{R}^n_{++} \to \mathbb{R}$ is convex if $f(x) = -\prod\limits_{i=1}^n x_i^{\alpha_i}$ if $\mathbf{1}^T\alpha = 1, \alpha \succeq 0$.

   Solution: Let's prove that $g(x) = -f(x)$ is concave: if g is concave than

$$\frac{\theta g(x) + (1-\theta)g(y)}{g(\theta x + (1-\theta)y)} \leqslant 1$$

   Now we can divide this in two parts:

$$\frac{g(x)}{g(\theta x + (1-\theta)y)} = \frac{\prod\limits_{i=1}^n x_i^{\alpha_i}}{\prod\limits_{i=1}^n (\theta x_i + (1-\theta)y_i)^{\alpha_i}} = \prod_{i=1}^n\left(\frac{x_i}{\theta x_i + (1-\theta)y_i}\right)^{\alpha_i} \leqslant \sum_{i=1}^n \frac{\alpha_i x_i}{\theta x_i + (1-\theta)y_i}$$

   and

$$\frac{g(y)}{g(\theta x + (1-\theta)y)} = \frac{\prod\limits_{i=1}^n y_i^{\alpha_i}}{\prod\limits_{i=1}^n (\theta x_i + (1-\theta)y_i)^{\alpha_i}} = \prod_{i=1}^n\left(\frac{y_i}{\theta x_i + (1-\theta)y_i}\right)^{\alpha_i} \leqslant \sum_{i=1}^n \frac{\alpha_i y_i}{\theta x_i + (1-\theta)y_i}$$

   Here we use AM-GM imequality.

   So now we can return to first inequality:

$$\frac{\theta g(x) + (1-\theta)g(y)}{g(\theta x + (1-\theta)y)} \leqslant \theta\sum_{i=1}^n \frac{\alpha_i x_i}{\theta x_i + (1-\theta)y_i} + (1-\theta)\sum_{i=1}^n \frac{\alpha_i y_i}{\theta x_i + (1-\theta)y_i} =$$

$$= \sum_{i=1}^n \frac{\alpha_i(\theta x_i + (1-\theta)y_i)}{\theta x_i + (1-\theta)y_i} = \sum_{i=1}^n \alpha_i = 1$$

   So if g is concave then f is convex.

8. Show that the maximum of a convex function f over the polyhedron $P = \text{conv}\{v_1, \dots, v_k\}$ is achieved at one of its vertices, i.e.,

$$\sup_{x \in P} f(x) = \max_{i=1,\dots,k} f(v_i).$$

A stronger statement is: the maximum of a convex function over a closed bounded convex set is achieved at an extreme point, i.e., a point in the set that is not a convex combination of any other points in the set. (you do not have to prove it). Hint: Assume the statement is false, and use Jensen's inequality. **Solution**: Lets's assume that $\exists x \in P : f(x) \geqslant \max_{i=1,\dots,k} f(v_i)$

So using Jensen's inequality we have:

$$f(x) = f\left(\sum_{i=1}^{k} \theta_i v_i\right) \leqslant \sum_{i=1}^{k} \theta_i f(v_i) \leqslant \sum_{i=1}^{k} \theta_i f(x) = f(x)$$

It means that

$$\sum_{i=1}^{k} \theta_i f(v_i) = f(x)$$

But we can always find $v^* = \max_{i=1,\dots,k} f(v_i)$ and

$$f(x) = \sum_{i=1}^{k} \theta_i f(v_i) \leqslant f(v^*)$$

So the statement that x is a maximizer f is false and v^* truly maximizer.

9. Show, that the two definitions of μ-strongly convex functions are equivalent: $f(x)$ is μ-strongly convex $\iff$ for any $x_1, x_2 \in S$ and $0 \leq \lambda \leq 1$ for some $\mu > 0$:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) - \frac{\mu}{2}\lambda(1-\lambda)\|x_1 - x_2\|^2$$

$f(x)$ is μ-strongly convex $\iff$ if there exists $\mu > 0$ such that the function $f(x) - \frac{\mu}{2}\|x\|^2$ is convex.

Solution: Firstly, let's show that second statement leads to first. So we know that $f(x) - \frac{\mu}{2}\|x\|^2$ is convex for some $\mu > 0$. We can use Jensen's inequality

$$f(\lambda x + (1-\lambda)y) - \frac{\mu}{2}\|\lambda x + (1-\lambda)y\|_2^2 \leqslant \lambda f(x) - \lambda\frac{\mu}{2}\|x\|_2^2 + (1-\lambda)f(y) - (1-\lambda)\frac{\mu}{2}\|y\|_2^2 \iff$$

$$\iff f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}(\lambda\|x\|_2^2 + (1-\lambda)\|y\|_2^2 - \|\lambda x + (1-\lambda)y\|_2^2)$$

Now we need to show that

$$\lambda\|x\|_2^2 + (1-\lambda)\|y\|_2^2 - \|\lambda x + (1-\lambda)y\|_2^2 = \lambda(1-\lambda)\|x-y\|_2^2$$

Firstly expand all brackets and simplify left side:

$$\lambda\|x\|_2^2 + (1-\lambda)\|y\|_2^2 - \|\lambda x + (1-\lambda)y\|_2^2 = \lambda\|x\|_2^2 + (1-\lambda)\|y\|_2^2 - \lambda^2\|x\|_2^2 - (1-\lambda)^2\|y\|_2^2 -$$

$$-2\lambda(1-\lambda)\langle x, y\rangle = \lambda(1-\lambda)\|x\|_2^2 + \lambda(1-\lambda)\|y\|_2^2 - 2\lambda(1-\lambda)\langle x, y\rangle$$

Now expand all brackets and simplify right side:

$$\lambda(1-\lambda)\|x-y\|_2^2 = \lambda(1-\lambda)\|x\|_2^2 + \lambda(1-\lambda)\|y\|_2^2 - 2\lambda(1-\lambda)\langle x, y\rangle$$

And they match perfectly!

Secondly, let's prove that first statement leads to second. We know that

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}\lambda(1-\lambda)\|x-y\|^2 =$$

$$= \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}\left(\lambda\|x\|_2^2 + (1-\lambda)\|y\|_2^2 - \|\lambda x + (1-\lambda)y\|_2^2\right) \iff$$

$$\iff f(\lambda x + (1-\lambda)y) - \frac{\mu}{2}\|\lambda x + (1-\lambda)y\|_2^2 \leqslant \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}\left(\lambda\|x\|_2^2 + (1-\lambda)\|y\|_2^2\right)$$

And attentive reader can see that it's Jensen's inequality for function $f(x) - \frac{\mu}{2}\|x\|_2^2$. It means that this function is convex for some $\mu > 0$.

Optimality conditions

1. Toy example

$$x^2 + 1 \to \min_{x \in \mathbb{R}}$$
$$\text{s.t. } (x-2)(x-4) \leq 0$$

   1. Give the feasible set, the optimal value, and the optimal solution.
   2. Plot the objective x^2+1 versus x. On the same plot, show the feasible set, optimal point, and value, and plot the Lagrangian $L(x,\mu)$ versus x for a few positive values of μ. Verify the lower bound property ($p^* \geq \inf_x L(x,\mu)$for $\mu \geq 0$). Derive and sketch the Lagrange dual function g.
   3. State the dual problem, and verify that it is a concave maximization problem. Find the dual optimal value and dual optimal solution μ^*. Does strong duality hold?
   4. Let $p^*(u)$ denote the optimal value of the problem

$$x^2 + 1 \to \min_{x \in \mathbb{R}}$$
$$\text{s.t. } (x-2)(x-4) \leq u$$

as a function of the parameter u. Plot $p^*(u)$. Verify that $\dfrac{dp^*(0)}{du} = -\mu^*$

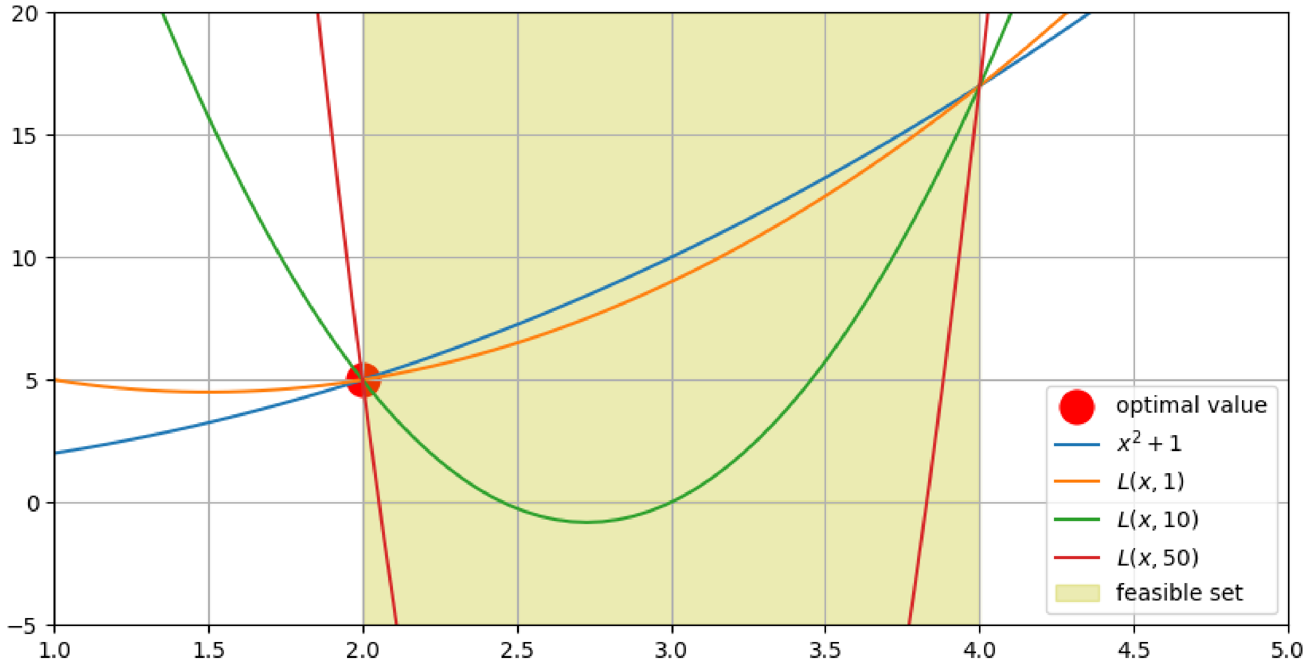
Solution:

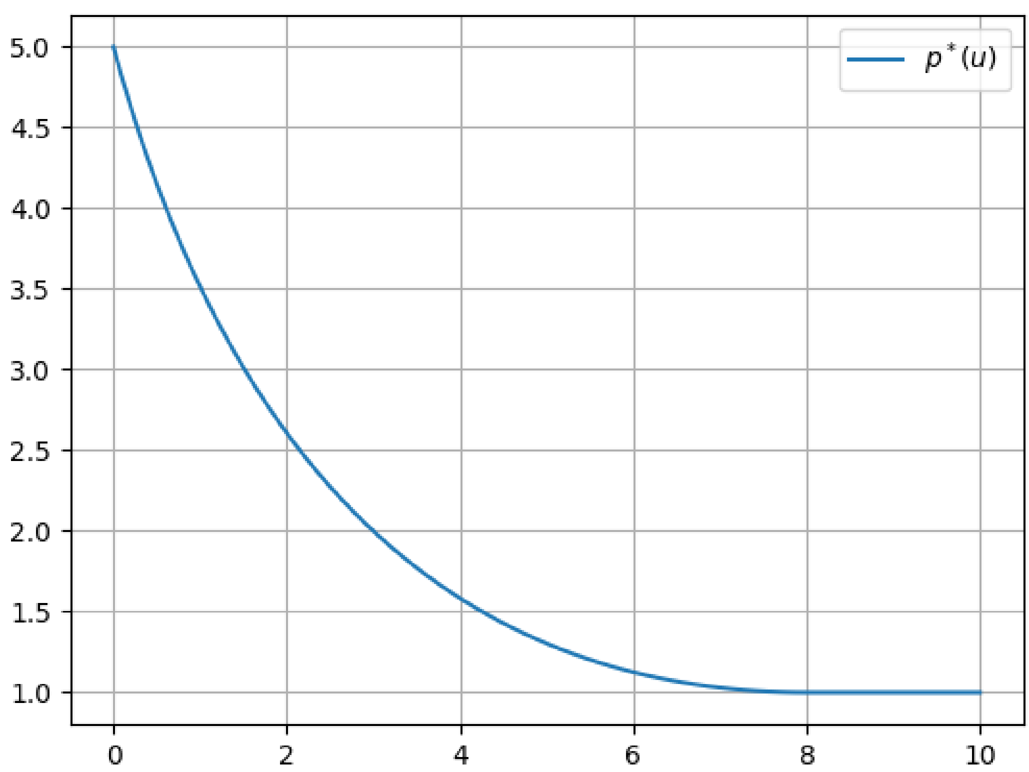
1. $(x-2)(x-4) \leqslant 0 \iff x \in [2,4]$ - the feasible set; $x^* = 2, p^* = 5$
2. $L(x,\mu) = x^2 + 1 + \mu(x-2)(x-4)$ $\nabla_x L(x,\mu) = 3x + 2\mu x - 6\mu = 0 \implies \hat{x} = \frac{3\mu}{1+\mu} \implies \inf_x L(x,\mu) = \frac{-\mu^2+9\mu+1}{1+\mu}$ $5 \geqslant \frac{-\mu^2+9\mu+1}{1+\mu} \iff 5 + 5\mu \geqslant -\mu^2 + 9\mu + 1 \iff 0 \leqslant (\mu-2)^2$- this is true.
3. $\frac{-\mu^2+9\mu+1}{1+\mu} \to \max_{\mu \geqslant 0} \implies \frac{\partial}{\partial \mu} = \frac{-\mu^2-2\mu+8}{(1+\mu)^2}, \frac{\partial^2}{\partial \mu^2} = \frac{-18}{(1+\mu)^3} < 0$ for $\mu \geqslant 0$ - it is a concave maximization problem $\implies \mu^* = 2$ cause $\mu \geqslant 0 \implies d^* = \frac{-4+18+1}{1+2} = 5$

So we can see that strong duality holds.

4. $(x-2)(x-4) \leqslant u \implies x^2 - 6x + 8 - u = 0 \implies x_{1,2} = 3 \pm \sqrt{1+u}$

$$p^*(u) = f(3-\sqrt{1+u})[3-\sqrt{1-u} \geqslant 0] + f(0)[3-\sqrt{1-u} < 0] =$$
$$= f(3-\sqrt{1+u})[u \leqslant 8] + f(0)[u > 8]$$
$$\frac{\partial p^*(0)}{\partial u} = 2 \cdot (3-\sqrt{1+0}) \cdot (-\frac{1}{2\sqrt{1+0}}) = -2 = -\mu^*$$





2. Derive the dual problem for the Ridge regression problem with $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, \lambda > 0$:

$$\frac{1}{2}\|y-b\|^2 + \frac{\lambda}{2}\|x\|^2 \to \min_{x\in\mathbb{R}^n, y\in\mathbb{R}^m}$$
$$\text{s.t. } y = Ax$$

Solution:

$$L(x,y,\mu) = \frac{1}{2}\|y-b\|^2 + \frac{\lambda}{2}\|x\|^2 + \mu^\top(y-Ax)\frac{\partial L}{\partial x} = \lambda x - A^\top\mu = 0 \implies x^* = \frac{1}{\lambda}A^\top\mu$$

$$\frac{\partial L}{\partial y} = y - b + \mu = 0 \implies y^* = b - \mu$$

So now we can check if x^* and y^* are the global minimizers:

$$\frac{\partial^2 L}{\partial x^2} = \lambda I > 0$$

$$\frac{\partial^2 L}{\partial y^2} = I > 0$$

$$g(\mu) = \inf_{x,y} L(x,y,\mu) = \frac{1}{2}\|-\mu\|^2 + \frac{\lambda}{2}\left\|\frac{1}{\lambda}A^\top\mu\right\|^2 + \mu^\top(b-\mu-\frac{1}{\lambda}AA^\top\mu) = \; = -\frac{1}{2}\|\mu\|^2 - \frac{1}{2\lambda}\|A^\top\mu\|^2 + \mu^\top b \to \max_{\mu}$$

3. Derive the dual problem for the support vector machine problem with $A \in \mathbb{R}^{m\times n}, \mathbf{1} \in \mathbb{R}^m \in \mathbb{R}^m, \lambda > 0$:

$$\langle \mathbf{1}, t\rangle + \frac{\lambda}{2}\|x\|^2 \to \min_{x\in\mathbb{R}^n, t\in\mathbb{R}^m}$$
$$\text{s.t. } Ax \succeq \mathbf{1} - t$$
$$t \succeq 0$$

Solution:

$$L(x,t,\mu_1,\mu_2) = \langle \mathbf{1}, t\rangle + \frac{\lambda}{2}\|x\|^2 - \mu_1^\top(Ax - \mathbf{1} + t) - \mu_2^\top t$$

$$\frac{\partial L}{\partial x} = \lambda x - A^\top\mu_1 = 0, \quad \frac{\partial^2 L}{\partial x^2} = \lambda > 0 \implies x^* = \frac{1}{\lambda}A^\top\mu_1$$

$$\frac{\partial L}{\partial t} = 1 - \mu_1 - \mu_2 = 0 \implies \mu_2 = 1 - \mu_1 \in [0,1], \mu_1 \in [0,1] \implies t^* = 0 \text{ or } t^* = \mathbf{1} - Ax$$

$$t^* = 0: \quad L(x^*, t^*, \mu_1, \mu_2) = \frac{\lambda}{2}\|x^*\|^2 - \mu_1^\top(Ax^* - \mathbf{1})$$

Another case has the same Lagrange value (pls belive me)

$$g(\mu_1) = \frac{\lambda}{2}\left\|\frac{1}{\lambda}A^\top\mu_1\right\|^2 - \langle\mu_1, \frac{1}{\lambda}AA^\top\mu_1 - \mathbf{1}\rangle = -\frac{1}{2\lambda}\|A^\top\mu_1\|^2 + \langle\mu_1, \mathbf{1}\rangle \to \max_{\mu_1\in[0,1]^m}$$

4. Give an explicit solution to the following LP.

$$
\begin{aligned}
&c^\top x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } &1^\top x = 1, \\
&x \succeq 0
\end{aligned}
$$

This problem can be considered the simplest portfolio optimization problem. **Solution**:

$$
L(x, \lambda, \mu) = c^\top x - \lambda^\top x + \mu(\mathbf{1}^\top x - 1)
$$

$$
g(\mu) = -\mu \to \max \quad \text{s.t. } \; c - \lambda + \mu\mathbf{1} = 0, \;\; \lambda \geqslant 0
$$

$$
\mu = \lambda_i - c_i \geqslant -\min_j c_j = \mu^*
$$

$$
p^* \geqslant d^* \implies x^* = (0 \ldots \underbrace{1}_{j} \ldots 0)
$$

5. Show, that the following problem has a unique solution and find it:

$$
\begin{aligned}
&\langle C^{-1}, X \rangle - \log\det X \to \min_{x \in \mathbb{R}^{n \times n}} \\
\text{s.t. } &\langle Xa, a \rangle \leq 1,
\end{aligned}
$$

where $C \in \mathbb{S}^n_{++}, a \in \mathbb{R}^n \neq 0$. The answer should not involve inversion of the matrix C. **Solution**:

$$
L(X, \lambda) = \langle C^{-1}, X \rangle - \log\det X + \lambda(\langle Xa, a \rangle - 1)
$$

$$
\nabla_x L = C^{-1} - X^{-1} + \lambda aa^\top = 0
$$

$$
\nabla_y L = \langle Xa, a \rangle - 1 = 0
$$

So now we can use Sherman-Morrison equality:

$$
C = (X^{-1} - \lambda aa^\top)^{-1} = X + \frac{\lambda X aa^\top X}{1 - \lambda}
$$

$$
a^\top C a = a^\top X a + \frac{\lambda}{1-\lambda} a^\top X a a^\top X a = \frac{1}{1-\lambda} \implies \lambda^* = 1 - \frac{1}{a^\top C a}
$$

$$
X = (C^{-1} + \lambda aa^\top)^{-1} = C\left(I - \frac{aa^\top C - \frac{aa^\top C}{a^\top C a}}{a^\top C a}\right)
$$

Notice that if $\lambda = 1$ then C is singular but, as you can see, $\lambda^* \neq 1$.

6. Give an explicit solution to the following QP.

$$
\begin{aligned}
&c^\top x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } &(x - x_c)^\top A (x - x_c) \leq 1,
\end{aligned}
$$

where $A \in \mathbb{S}^n_{++}, c \neq 0, x_c \in \mathbb{R}^n$ **Solution**:

$$
L(x, \lambda) = c^\top x + \lambda((x - x_c)^\top A (x - x_c) - 1)
$$

$$
\nabla_x L = c + 2\lambda A (x - x_c) = 0 \implies x = x_c - \frac{A^{-1} c}{2\lambda}
$$

$$\nabla_\lambda L = (x - x_c)^\top A(x - x_c) - 1 = 0$$

$$\left(x_c - \frac{A^{-1}c}{2\lambda} - x_c\right)^\top A \left(x_c - \frac{A^{-1}c}{2\lambda} - x_c\right) = \frac{1}{4\lambda^2} c^\top A^{-1} c = 1 \implies \lambda^* = \frac{1}{2}\sqrt{c^\top A^{-1} c}$$

$$x^* = x_c - \frac{A^{-1}c}{\sqrt{c^\top A^{-1} c}}$$

7. Consider the equality-constrained least-squares problem

$$\|Ax - b\|_2^2 \to \min_{x \in \mathbb{R}^n}$$
$$\text{s.t. } Cx = d,$$

where $A \in \mathbb{R}^{m \times n}$ with $\textbf{rank}A = n$, and $C \in \mathbb{R}^{k \times n}$ with $\textbf{rank}C = k$. Give the KKT conditions, and derive expressions for the primal solution x^* and the dual solution λ^*. **Solution**:

$$L(x, \lambda) = \|Ax - b\|_2^2 + \lambda^\top (Cx - d)$$

The KKT conditions are

$$\nabla_x L = 2A^\top (Ax - b) + C^\top \lambda = 0$$
$$\nabla_\lambda L = Cx - d = 0$$

These conditions satisfy the Slater's condition.

$$x = \frac{1}{2}(A^\top A)^{-1}(2A^\top b - C^\top \lambda)$$

$$Cx - d = C\frac{1}{2}(A^\top A)^{-1}(2A^\top b - C^\top \lambda) - d = C(A^\top A)^{-1}A^\top b - \frac{1}{2}C(A^\top A)^{-1}C^\top \lambda - d = 0$$
$$\lambda^* = 2(C(A^\top A)^{-1}C^\top)^{-1}(C(A^\top A)^{-1}A^\top b - d)$$
$$x^* = (A^\top A)^{-1}A^\top b - (A^\top A)^{-1}C^\top (C(A^\top A)^{-1}C^\top)^{-1}(C(A^\top A)^{-1}A^\top b - d)$$

8. Derive the KKT conditions for the problem

$$\textbf{tr } X - \log\det X \to \min_{X \in \mathbb{S}^n_{++}}$$
$$\text{s.t. } Xs = y,$$

where $y \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ are given with $y^\top s = 1$. Verify that the optimal solution is given by

$$X^* = I + yy^\top - \frac{1}{s^\top s} ss^\top$$

Solution:

$$L(x, \lambda) = \textbf{tr } X - \log\det X + \lambda^\top (Xs - y)$$

The KKT conditions are

$$\nabla_x L = I - X^{-1} + \lambda s^\top = 0$$
$$\nabla_\lambda L = Xs - y = 0$$

So we have

$$X = (I + \lambda s^\top)^{-1} \implies Xs = (I + \lambda s^\top)^{-1} s = y \implies \lambda = s - y$$

$$X = (I + \lambda s^\top)^{-1} = I - \frac{\lambda s^\top}{1 + s^\top \lambda} = I + \frac{y s^\top}{s^\top s} - \frac{s s^\top}{s^\top s}$$

Our answer is close to the given optimal solution, so we have to prove that $y = \frac{s}{s^\top s}$

$$y = \frac{s}{s^\top s} \iff s = s^\top s y \iff s^\top s = s^\top \underbrace{s^\top s}_{\in \mathbb{R}} y = s^\top s \underbrace{s^\top y}_{=1} = s^\top s$$

So our answer perfectly matches with the given one.

9. Supporting hyperplane interpretation of KKT conditions. Consider a convex problem with no equality constraints

$$\begin{split} & f_0(x) \to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } & f_i(x) \leq 0, \quad i = [1, m] \end{split}$$

Assume, that $\exists x^* \in \mathbb{R}^n, \mu^* \in \mathbb{R}^m$ satisfy the KKT conditions

$$\begin{split} & \nabla_x L(x^*, \mu^*) = \nabla f_0(x^*) + \sum_{i=1}^m \mu_i^* \nabla f_i(x^*) = 0 \\ & \mu_i^* \geq 0, \quad i = [1, m] \\ & \mu_i^* f_i(x^*) = 0, \quad i = [1, m] \\ & f_i(x^*) \leq 0, \quad i = [1, m] \end{split}$$

Show that

$$\nabla f_0(x^*)^\top (x - x^*) \geq 0$$

for all feasible x. In other words, the KKT conditions imply the simple optimality criterion or $\nabla f_0(x^*)$ defines a supporting hyperplane to the feasible set at x^*. **Solution**: If conditions inactive this means that $\mu = 0$ and this leads to $\nabla f_i = 0$. So

$$\nabla f_0(x^*)^\top (x - x^*) = 0(x - x^*) = 0 \geqslant 0$$

Else we can find at least one active condition $\implies f_i(x^*) = 0$. Now we have 2 cases:

1) If feasible x for $f_i(x)$ greater that x^* then $f_0(x) < f_0(x^*)$. In this case $\nabla f_0(x^*) \geqslant 0$ and finally

$$f_0(x^*)^\top (x - x^*) \geqslant 0$$

2) If feasible x for $f_i(x)$ less that x^* then $f_0(x) > f_0(x^*)$. In this case $\nabla f_0(x^*) \leqslant 0$ and finally

$$f_0(x^*)^\top (x - x^*) \geqslant 0$$

10. A penalty method for equality constraints. We consider the problem of minimization

$$\begin{split} & f_0(x) \to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } & Ax = b, \end{split}$$

where $f_0(x) : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable, and $A \in \mathbb{R}^{m \times n}$ with $\textbf{rank} A = m$. In a quadratic penalty method, we form an auxiliary function

$$\phi(x) = f_0(x) + \alpha \|Ax - b\|_2^2,$$

where $\alpha > 0$ is a parameter. This auxiliary function consists of the objective plus the penalty term $\alpha\|Ax-b\|_2^2$. The idea is that a minimizer of the auxiliary function, $\tilde{x}$, should be an approximate solution to the original problem. Intuition suggests that the larger the penalty weight α, the better the approximation $\tilde{x}$ to a solution of the original problem. Suppose $\tilde{x}$ is a minimizer of $\phi(x)$. Show how to find, from $\tilde{x}$, a dual feasible point for the original problem. Find the corresponding lower bound on the optimal value of the original problem. **Solution**: First things first let's find optimal α^*

$$
\nabla_x \phi(\tilde{x}) = \nabla_x f_0(\tilde{x}) + 2\alpha A^\top(A\tilde{x} - b) = 0
$$

$$
\begin{gathered}
2\alpha A^\top(A\tilde{x} - b) = -\nabla_x f_0(\tilde{x}) \\
2\alpha(A^\top(A\tilde{x} - b))^\top A^\top(A\tilde{x} - b) = -(A^\top(A\tilde{x} - b))^\top \nabla_x f_0(\tilde{x}) \\
\alpha = -\frac{1}{2}\Big((A^\top(A\tilde{x} - b))^\top A^\top(A\tilde{x} - b)\Big)^{-1}(A^\top(A\tilde{x} - b))^\top \nabla_x f_0(\tilde{x})
\end{gathered}
$$

We can't say anything about right part but $\alpha > 0$, so let's

$$
\alpha^* = \max\left(\varepsilon, -\frac{1}{2}\Big((A^\top(A\tilde{x} - b))^\top A^\top(A\tilde{x} - b)\Big)^{-1}(A^\top(A\tilde{x} - b))^\top \nabla_x f_0(\tilde{x})\right)
$$

Now about lower bound: we know that

$$
f_0(x) \geqslant \inf L(x, \lambda)
$$

$$
\begin{gathered}
L(x, \lambda) = f_0(x) + \lambda^\top(Ax - b) \\
\nabla_x L = \nabla_x f_0(x) + A^\top\lambda = 0 \\
A\nabla_x f_0(x) = -AA^\top\lambda \\
\lambda = -(AA^\top)^{-1}A\nabla_x f_0(x) = 2\alpha^*(AA^\top)^{-1}AA^\top(A\tilde{x} - b) = 2\alpha^*(A\tilde{x} - b)
\end{gathered}
$$

And finally

$$
f_0(x) \geqslant L(\tilde{x}, \lambda^*) = f_0(\tilde{x}) + 2\alpha^*(A\tilde{x} - b)^\top(A\tilde{x} - b) = f_0(\tilde{x}) + 2\alpha^*\|A\tilde{x} - b\|^2
$$