FEDERAL STATE AUTONOMOUS
EDUCATIONAL INSTITUTION OF HIGHER EDUCATION
MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY
(STATE UNIVERSITY)
PHYSTECH-SCHOOL OF APPLIED MATHEMATICS AND COMPUTER SCIENCE

Homework.

Optimization

3-rd year student, group B05-003
Kreinin Matvei

Moscow, 2022

Contents

1 Matrix calculus **3**
1.1 Problem № 1 . . . 3
1.2 Problem № 2 . . . 3
1.3 Problem № 3 . . . 4
1.4 Problem № 4 . . . 5
1.5 Problem № 5 . . . 5
1.6 Problem № 6 . . . 6

2 Automatic differentiation **7**

3 Convex sets **14**
3.1 Problem №1 . . . 14
3.2 Problem №2 . . . 15
3.3 Problem №3 . . . 15
3.4 Problem №4 . . . 15
3.5 Problem №5 . . . 16

4 Convex functions **18**
4.1 Problem №1 . . . 18
4.2 Problem №2 . . . 18
4.3 Problem №3 . . . 19
4.4 Problem №4 . . . 20
4.5 Problem №5 . . . 21
4.6 Problem №6 . . . 21
4.7 Problem №7 . . . 21

5 Conjugate sets **23**
5.1 Problem №1 . . . 23
5.2 Problem №2 . . . 25
5.3 Problem №3 . . . 25
5.4 Problem №4 . . . 25
5.5 Problem №5 . . . 26

6 Conjugate functions **29**
6.1 Problem №1 . . . 29
6.2 Problem №2 . . . 29
6.3 Problem №3 . . . 29
6.4 Problem №4 . . . 30
6.5 Problem №5 . . . 31

7 Subgradient and subdifferential **32**
7.1 Problem № 1 32
7.2 Problem №2 32
7.3 Problem №3 33
7.4 Problem №4 33
7.5 Problem №5 33

8 General optimization problem **35**
8.1 Problem №1 35
8.2 Problem №2 35
8.3 Problem №3 36
8.4 Problem №4 36
8.5 Problem №5 37
8.6 Problem №6 39
8.7 Problem №7 39
8.8 Problem №8 39
8.9 Problem №9 40
8.10 Problem №10 42

9 Duality **43**
9.1 Problem №1 43
9.2 Problem №2 43
9.3 Problem №3 45
9.4 Problem №4 46

10 Linear Programming **48**
10.1 Problem №1 48
10.2 Problem №2 52
10.3 Problem №3 53
10.4 Problem №4 55

11 References **58**

1 Matrix calculus

1.1 Problem № 1

Find the gradient $\nabla f(x)$ and hessian $f^{''}(x)$, if $f(x) = \frac{1}{2}\|Ax - b\|_2^2$

Solution:

$$f(x) = \frac{1}{2}\langle Ax - b, Ax - b\rangle = \frac{1}{2}\langle Adx, Ax - b\rangle + \frac{1}{2}\langle Ax - b, Adx\rangle$$

$$f(x) = \frac{1}{2}\langle Ax - b, Adx\rangle + \frac{1}{2}\langle Ax - b, Adx\rangle = \langle Ax - b, Adx\rangle = \langle A^T(Ax - b), dx\rangle$$

$$\nabla f(x) = A^T(Ax - b)$$

$$df(x) = \langle A^T(Ax - b), dx\rangle$$

$$d^2f(x) = \langle d(A^T(Ax_2 - b), dx_1\rangle = \langle A^TAdx_2, dx_1\rangle = \langle dx_1, A^TAdx_2\rangle$$

$$d^2f(x) = \langle A^TAdx_1, dx_2\rangle$$

Answer: $\nabla f(x) = A^T(Ax - b)$, $f^{''}(x) = A^TA$

1.2 Problem № 2

Find gradient and hessian of $f : \mathbb{R}^n \to \mathbb{R}$, if:

$f(x) = \log\left(\sum\limits_{i=1}^{m} exp(a_i^Tx + b_i)\right), a_1, ..., a_m \in \mathbb{R}^n; b_1, ..., b_m \in \mathbb{R}$

Solution:

$$df(x) = \frac{d\left(\sum\limits_{i=1}^{m} exp(a_i^Tx + b_i)\right)}{\sum\limits_{i=1}^{m} exp(a_i^Tx + b_i)} = \frac{\sum\limits_{i=1}^{m} exp(a_i^Tx + b_i)a_i^Tdx}{\sum\limits_{i=1}^{m} exp(a_i^Tx + b_i)} = \frac{\langle\sum\limits_{i=1}^{m} exp(a_i^Tx + b_i)a_i^T, dx\rangle}{\sum\limits_{i=1}^{m} exp(a_i^Tx + b_i)}$$

$$\nabla f(x) = \frac{\sum\limits_{i=1}^{m} exp(a_i^Tx + b_i)a_i^T}{\sum\limits_{i=1}^{m} exp(a_i^Tx + b_i)}$$

$$d^2f(x) = \langle d\left(\frac{\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)a_i}{\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)}\right), dx_1\rangle$$

$$d^2f(x) = \langle\left(\frac{\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)a_i a_i^T}{\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)} + (\frac{\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)a_i a_i^T}{\left(\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)\right)^2}\right) dx_2, dx_1\rangle$$

$$d^2f(x) = \langle dx_1, \left(\frac{\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)a_i a_i^T}{\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)} + (\frac{\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)a_i a_i^T}{\left(\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)\right)^2}\right) dx_2\rangle$$

$$d^2f(x) = \langle\left(\frac{\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)a_i^T a_i}{\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)} + (\frac{\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)a_i^T a_i}{\left(\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)\right)^2}\right) dx_1, dx_2\rangle$$

Answer: $\nabla f(x) = \frac{\sum\limits_{i=1}^{m} exp(a_i^T x + b_i)a_i^T}{\sum\limits_{i=1}^{m} exp(a_i^T x + b_i)}$; $f''(x) = \left(\frac{\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)a_i^T a_i}{\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)} + (\frac{\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)a_i^T a_i}{\left(\sum\limits_{i=1}^{m} exp(a_i^T x_2 + b_i)\right)^2}\right)$

1.3 Problem № 3

Calculate the derivatives of the loss function with respect to parameters $\frac{\partial L}{\partial W}, \frac{\partial L}{\partial b}$ for the single object $x_i (or, n = 1)$

Solution:

$$L = \frac{1}{n}\sum_{i=1}^{n}\|y_i - \tilde{y}\|^2 = \frac{1}{n}\sum_{i=1}^{n}\langle y_i - \tilde{y}, y_i - \tilde{y}\rangle = \frac{1}{n}\sum_{i=1}^{n}\langle y_i - Wx_i - b, y_i - Wx_i - b\rangle$$

$$dL(dW) = \frac{1}{n}\sum_{i=1}^{n}\langle y_i - Wx_i - b, -dWx_i,\rangle + \langle y_i - Wx_i - b, -dWx_i\rangle$$

$$dL(dW) = \frac{2}{n}\sum_{i=1}^{n}\langle -dWx_i, y_i - Wx_i - b\rangle = -\frac{2}{n}\sum_{i=1}^{n}\langle (y_i - Wx_i - b)x_i^T, dW\rangle$$

$$dL(db) = \frac{1}{n}\sum_{i=1}^{n}\langle -db, y_i - Wx_i - b\rangle + \langle y_i - Wx_i - b, -db\rangle = -\frac{2}{n}\sum_{i=1}^{n}\langle y_i - Wx_i - b, db\rangle$$

Answer: $\frac{\partial L}{\partial W} = -\frac{2}{n}\sum\limits_{i=1}^{n}(y_i - Wx_i - b)x_i^T$; $\frac{\partial L}{\partial b} = -\frac{2}{n}\sum\limits_{i=1}^{n} y_i - Wx_i - b$

1.4 Problem № 4

Calculate:

$$\frac{\partial}{\partial X}\sum \text{eig(X)}, \frac{\partial}{\partial X}\prod \text{eig(X)}, \frac{\partial}{\partial X}tr(X), \frac{\partial}{\partial X}\text{det(X)}$$

Solution:

$$\frac{\partial}{\partial X}\sum \text{eig(X)} = \frac{\partial}{\partial X}tr(X)$$

$$d(tr(X)) = tr(dX) = tr(I^T, dX) = \langle I, dX\rangle$$

$$\frac{\partial}{\partial X}\prod \text{eig(X)} = \frac{\partial}{\partial X}\text{det(X)}$$

$$det(X) = \sum_{i=1}^{n} x_{ij}M_{ij}; \frac{\partial X}{\partial x_{ij}} = \frac{\partial \sum\limits_{i=1}^{n} x_{ij}M_{ij}}{\partial x_{ij}} = M_{ij}$$

Because $x_{ij}^{-1} = \frac{M_{ji}}{detX}$, then

$$\frac{\partial(det(X))}{\partial X} = det(X)X^{-T}$$

Answer: $\frac{\partial}{\partial X}\sum \text{eig(X)} = \frac{\partial}{\partial X}tr(X) = I$; $\frac{\partial}{\partial X}\prod \text{eig(X)} = \frac{\partial}{\partial X}\text{det(X)} = det(X)X^{-T}$

1.5 Problem № 5

Calculate the first and the second derivative of the following function: $f : S \to \mathbb{R}$
$f(t) = det(A - tI_n)$, where $A \in \mathbb{R}^{n\dot{n}}, S := \{t \in \mathbb{R} : det(A - tI_n) \neq 0\}$
Solution:

$$df(t) = det(A - t\cdot I)\langle(A - t\cdot I)^{-T}, -Idt\rangle = -det(A - t\cdot I)\langle(A - t\cdot I)^{-T}, -Idt\rangle$$

$$df(t) = -f(t)\cdot tr\left((A - t\cdot I)^{-1}\right)dt$$

Okay, let's try to calculate second derivative of that nice function!

$$d^2f(t) = -d\left(f(t)\cdot tr\left((A-t\cdot I)^{-1}\right)dt_1\right)$$

$$d^2f(t) = -\nabla f(t)\cdot tr\left((A-t\cdot I)^{-1}\right)dt_2\cdot dt_1 - f(t)\langle I, -(A-t\cdot I)^{-1}(-Idt_2)(A-t\cdot I)^{-1}\rangle dt_1$$

And then we get:

$$d^2f(t) = -\left(\nabla f(t)\cdot tr\left((A-t\cdot I)^{-1}\right) + f(t)\cdot tr\left(((A-t\cdot I)^{-2})^T\right)\right)\cdot dt_1\cdot dt_2$$

Answer: $\nabla f(t) = -f(t)\cdot tr\left((A-t\cdot I)^{-1}\right)$
$f''(t) = -\left(\nabla f(t)\cdot tr\left((A-t\cdot I)^{-1}\right) + f(t)\cdot tr\left(((A-t\cdot I)^{-2})^T\right)\right)$

1.6 Problem № 6

Find the gradient $\nabla f(x)$, if $f(x) = tr(AX^2BX^{-T})$.
Solution:

$$df(X) = d(tr(AX^2BX^{-T}) = \langle I, d(AX^2BX^{-T})\rangle$$

$$df(X) = \langle I, A(XdX + dXX)BX^{-T} - AX^2BX^{-T}dX^TX^{-T}\rangle$$

$$df(X) = \langle (BX^{-T}AX)^T, dX\rangle + \langle (XBX^{-T}A)^T, dX\rangle + \langle (X^{-T}AX^2BX^{-T})^T, dX^T\rangle$$

$$df(X) = \langle X^TA^TX^{-1}B^T, dX\rangle + \langle A^TX^{-1}B^TX^T, dX\rangle - \langle X^{-1}B^TX^TX^TA^TX^{-1}, dX^T\rangle$$

$$df(X) = \langle X^TA^TX^{-1}B^T + A^TX^{-1}B^TX^T, dX\rangle - \langle (X^{-1}B^TX^TX^TA^TX^{-1})^T, dX\rangle$$

$$df(X) = \langle X^TA^TX^{-1}B^T + A^TX^{-1}B^TX^T - X^{-T}AXXBX^{-T}, dX\rangle$$

Answer: $\nabla f(x) = X^TA^TX^{-1}B^T + A^TX^{-1}B^TX^T - X^{-T}AXXBX^{-T}$

2 Automatic differentiation

In [155]:

```
import jax
import numpy

from numpy.linalg import inv
from jax import numpy as jnp
from jax import grad
```

Problem №1

You will work with the following function for exercise, $f(x, y) = e^{-(sin(x)+cos(y))^2}$

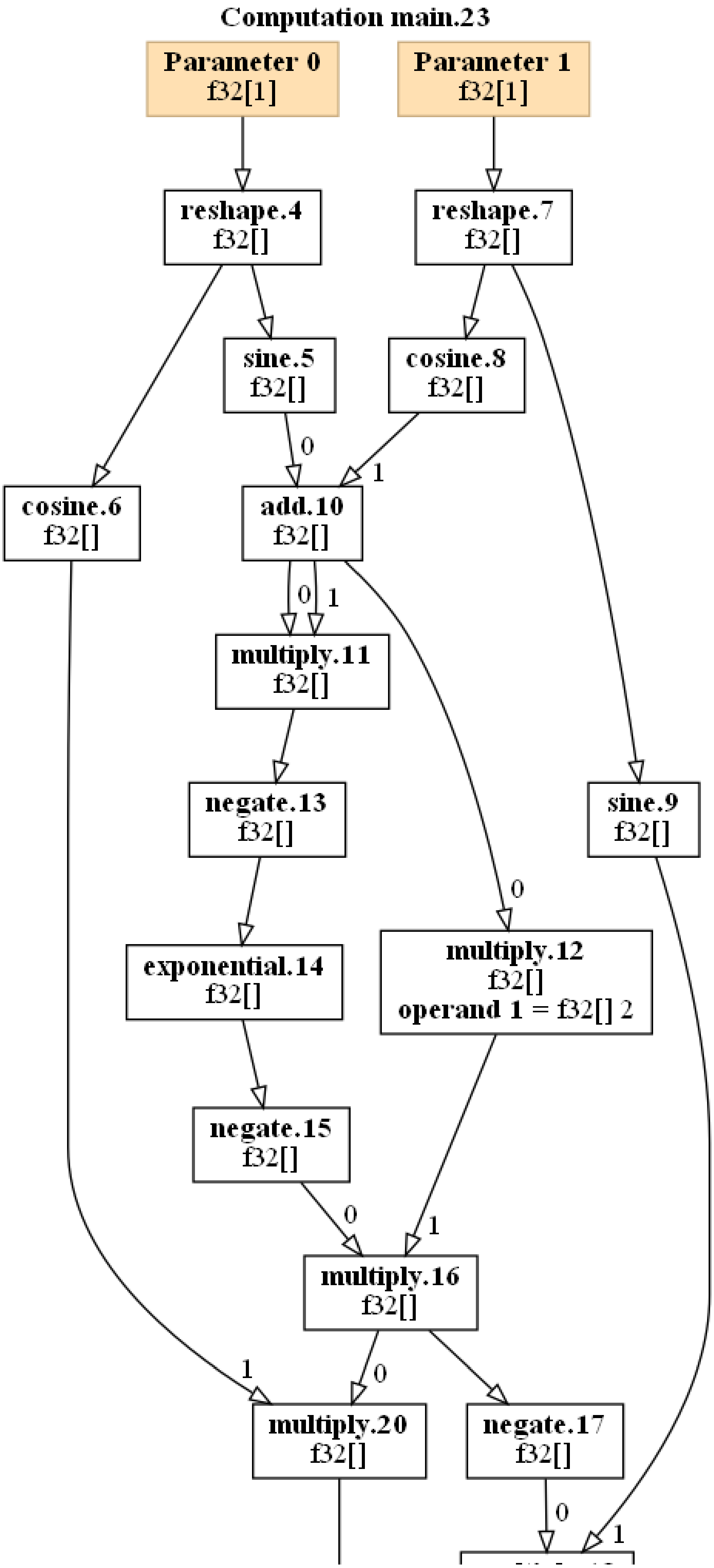
Draw the computational graph for the function. Note, that it should contain only primitive operations - you need to do it automatically.

In [156]:

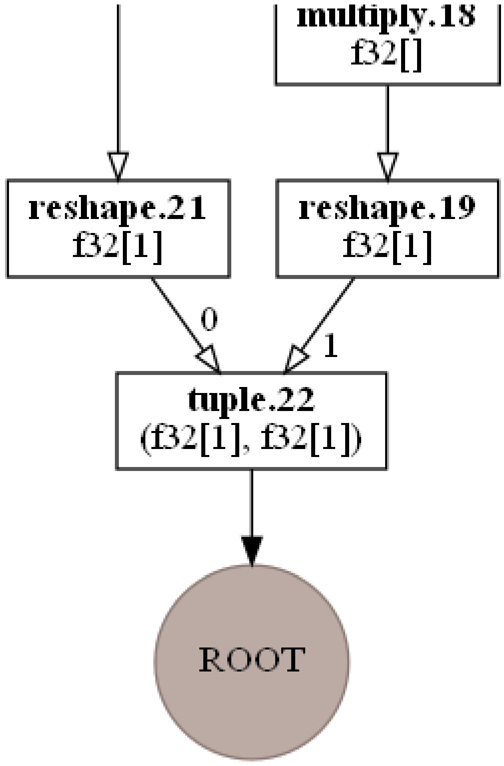
```
#Function of first problem
def func_p1(x, y):
  return jnp.exp(- jnp.power((jnp.sin(x[0]) + jnp.cos(y[0])), 2))

def dfunc_p1(x, y):
  return grad(func_p1, argnums=(0, 1))(x, y)
```

In [157]:

```
z=jax.xla_computation(dfunc_p1)(numpy.random.rand(1), numpy.random.rand(1))

with open("t1.txt", "w") as f:
  f.write(z.as_hlo_text())

with open("t1.dot", "w") as f:
  f.write(z.as_hlo_dot_graph())
```



Computation main.23
Parameter 0
f32[1]
Parameter 1
f32[1]
reshape.4
f32[]
reshape.7
f32[]
sine.5
f32[]
cosine.8
f32[]
cosine.6
f32[]
add.10
f32[]
multiply.11
f32[]
negate.13
f32[]
sine.9
f32[]
exponential.14
f32[]
multiply.12
f32[]
operand 1 = f32[] 2
negate.15
f32[]
multiply.16
f32[]
multiply.20
f32[]
negate.17
f32[]



Problem №2

Compare analytic and autograd approach for the hessian of: $f(x) = \frac{1}{2}x^T Ax + b^T x + c$

In [158]:

```
from jax import jacfwd, jacrev
```

In [159]:

```
A = numpy.random.rand(100, 100)
b = numpy.random.rand(100)
c = 1

def func_p2(x):
 return 0.5 * x.T @ A @ x + b @ x + c

def hessian(f):
  return jax.jacfwd(jax.grad(f))

def d2func_p2(x):
  return hessian(func_p2)(x)

hessian_auto2 = d2func_p2(numpy.random.rand(100))
hessian_anal2 = (A + A.T) / 2
```

Difference between autograde and analytical solution:

In [160]:

```
numpy.linalg.norm(hessian_anal2 - hessian_auto2)
```

Out[160]:

```
2.1366172e-06
```

Cringe moment for visualising it

In [161]:

```
z = jax.xla_computation(d2func_p2)(numpy.random.rand(100))

with open("t2.txt", "w") as f:
   f.write(z.as_hlo_text())

with open("t2.dot", "w") as f:
  f.write(z.as_hlo_dot_graph())
```

Problem №3

Suppose we have the following function $f(x) = \frac{1}{2}||x||^2$, select a random point $x_0 \in \mathbb{B}^{1000} = \{0 \leq x_i \leq 1 | \forall i\}$. Consider 10 steps of the gradient descent starting from the point x_0:

$x_{k+1} = x_k - \alpha_k \nabla f(x_k).$

Your goal in this problem to write the function, that takes 10 scalar values α_i and return the result of the gradient descent on function $L = f(x_{10})$. And optimize this function using gradient descent on $\alpha \in \mathbb{R}^{10}$. Suppose, $\alpha_0 = 1$.

$\alpha_{k+1} = \alpha_k - \beta \cdot \frac{\partial L}{\partial \alpha}$

Choose any β and the number of steps your need. Describe obtained results.

In [162]:

```
import numpy as np
```

In [163]:

```
def dldalpha_p3(x, alpha):
    N = len(alpha) - 1
    dl = np.zeros(N + 1)
    cur_alpha = 1 - alpha[N]

    for k in range(N):
        dl[N - k ] = -x[N - k] @ x[N - k - 1] * cur_alpha
        cur_alpha *= (1-alpha[N - k - 1])

    return dl

def my_grad_p3(alpha, iterations):
    x = np.array([np.random.rand(1000) for i in range(len(alpha) + 1)])
    beta = 3e-3

    for l in range(iterations):
        for i in range(len(alpha)):
            x[i+1] =  x[i] - alpha[i] * x[i]

        alpha = alpha - beta * dldalpha_p3(x, alpha)

    return x
```

In [164]:

```
alpha = np.array([float(np.random.rand(1)) for i in range(10)])
x = my_grad_p3(alpha, 3)
print(x[-1])
# The method converges very quickly, despite the number of iterations (3 or 30),
# and then e-5 is already a machine zero hinders.
# Having tested it several times, everything strongly depends on the initial alpha.
# The closer to one, the better
```

```
 4.73998639e-04 1.09616949e-04 1.41240217e-04 4.28934702e-05
 1.21931998e-04 2.55045152e-04 2.57784649e-04 6.62678159e-05
 3.57227559e-04 3.54347994e-04 5.05101070e-04 4.57211362e-04
 2.12416394e-05 2.65531852e-04 2.75066628e-04 4.96563818e-04
 2.42761458e-04 6.84828172e-05 3.54887245e-04 1.75792293e-04
 4.96970409e-04 3.40873064e-04 1.14075589e-04 2.57406378e-04
 4.46789599e-04 2.32420815e-04 2.56337498e-04 2.01395761e-04
 8.06003241e-05 1.72717746e-04 9.82050641e-05 2.73584191e-04
 8.09531208e-05 2.11171733e-04 1.74160573e-04 7.33927375e-05
 4.80571989e-04 3.52904366e-04 4.45467706e-04 4.53448745e-05
 2.54705059e-04 2.70298146e-04 3.09885902e-04 1.72345172e-04
 4.60182349e-04 5.03076812e-04 1.58222722e-04 1.56601982e-04
 3.01523796e-04 3.47146878e-04 5.10606641e-04 3.21960333e-04
 2.58708699e-04 5.11791452e-04 4.37511632e-04 3.89106132e-05
 5.16222657e-05 5.31700141e-04 3.76305500e-04 3.32173200e-04
 6.28505105e-05 1.91547715e-05 1.46589186e-04 6.56351555e-05
 2.96961921e-05 3.34641570e-04 3.21975249e-04 1.70076498e-04
 3.23673641e-05 5.77914347e-05 2.63653661e-04 4.60466515e-04
 1.74001251e-05 2.56137355e-04 2.19101837e-04 2.08307824e-04]
```

Problem №4

Compare analytic and autograd approach for the gradient of: $f(X) = -log(det(X))$

Analytical gradient: $df = -\frac{1}{det(X)} \cdot det(X)\langle X^{-T}, dX\rangle$

$df = -\langle X^{-T}, dX\rangle$

$\nabla f = -X^{-T}$

In [165]:

```
X = numpy.random.rand(100, 100)

func_p4 = lambda X: -jnp.log(jnp.linalg.det(X))
dfunc_p4 = lambda X: grad(func_p4)(X)

grad_auto4 = dfunc_p4(X)
grad_anal4 = -(inv(X)).T

print("Difference between analytical and autograde methods:",
      numpy.linalg.norm(grad_auto4 - grad_anal4))
```

```
Difference between analytical and autograde methods: 0.00013053074
```

Problem №5

Compare analytic and autograd approach for the gradient and hessian of: $f(x) = x^T x x^T x$

In [166]:

```
def func_p5(x):
  return jnp.dot(x.T, x)* jnp.dot(x.T, x)

def dfunc_p5(x):
  return grad(func_p5)(x)

def d2func_p5(x):
  return hessian(func_p5)(x)
```

Analytical gradient: $df = 4\langle x, x\rangle \cdot \langle x, dx\rangle$

$\nabla f = 4\langle x, x\rangle \cdot x$

Analytical hessian: $d^2 f = 4 \cdot \left(\langle dx_2, x\rangle \cdot \langle x, dx_1\rangle + \langle x, dx_2\rangle \cdot \langle x, dx_1\rangle + \langle x, x\rangle \cdot \langle dx_2, dx_1\rangle\right)$

$$d^2 f = 8 \cdot (x, dx_1)(x, dx_2) + 4(x, x)(dx_2, dx_1) = (8x(x, dx_2), dx_1) + (4(x, x)dx_2, dx_1) =$$
$$= (8xx^T dx_2, dx_1) + (4(x, x)dx_2, dx_1) = ((8xx^T + 4(x, x)I)dx_2, dx_1)$$

$hessian(f) = 8xx^T + 4(x, x)I$

In [167]:

```
x = numpy.random.rand(10)
grad_auto5 = grad(func_p5)(x)
grad_anal5 = 4 * jnp.dot(x, x) * x

print("Difference between analytic and auto gradient",
      numpy.linalg.norm(grad_auto5 - grad_anal5))
```

```
Difference between analytic and auto gradient 0.0
```

In [168]:

```
hessian_auto5 = d2func_p5(x)
hessian_anal5 = 8*jnp.outer(x, x.T) + 4 * x @ x * jnp.eye(len(x))

print("Difference between analytic and auto hessian:",
      numpy.linalg.norm(hessian_auto5 - hessian_anal5))
```

```
Difference between analytic and auto hessian: 6.031566e-06
```

3 Convex sets

3.1 Problem №1

Show that the convex hull of the **S** set is the intersection of all convex sets containing **S**

<u>**Solution:**</u>

Firstly, we will prove that if A is a convex set, then any convex combination $x_1, .., x_n \in A$ will belong to A.

For $n = 1$ it's trivial.

Assume that this is true for any convex combination of n-1 points. Point $x = \sum\limits_{j=1}^{n} \alpha_j x_j$, $\sum\limits_{j=1}^{n} \alpha_j = 1$, $\alpha_1, ..., \alpha_n \in \mathbb{R}$ and $n > 1$. Between $\alpha_1, ..., \alpha_n$ we can find α that will not be equal to 1. Without limiting the generality we consider that $\alpha \neq 1$.

$$\overline{\alpha_j} = \frac{\alpha_j}{1 - \alpha_1},\ j = 2, ..., n$$

Because $\sum\limits_{j=2}^{n} \overline{\alpha_j} = 1$ and from induction we can get that $\overline{x} = \sum\limits_{j=1}^{n} \overline{\alpha_j} x_j \in A$ Then from convex of A we get:

$$x = \sum_{j=1}^{n} \alpha_j x_j = \alpha_1 x_1 + (1 - \alpha_1) \sum_{j=1}^{n} \frac{\alpha_j}{1 - \alpha_1} x_j = \alpha_1 x_1 + (1 - \alpha_1)\overline{x} \in A$$

From the proven it follows that if some convex set contains A, then it contains any convex combination of points from A, which means it contains convex hull of A. Let's show that convex hull of A is a convex set, in that case it coincides with the intersection of all convex sets containing A. Let's take random points from convex hull A.

$$x = \sum_{j=1}^{n} \alpha_j x_j,\ y = \sum_{j=1}^{m} \beta_j y_j$$

For any $\alpha \in [0, 1]$, we get:

$$(1 - \alpha)x + \alpha y = \sum_{j=1}^{n} (1 - \alpha)\alpha_j x_j + \sum_{j=1}^{m} \alpha \beta_j y_j$$

$$\sum_{j=1}^{n} (1 - \alpha)\alpha_j + \sum_{j=1}^{m} \alpha \beta_j = (1 - \alpha) + \alpha = 1$$

We get a convex combination of points $x_1, ..., x_n, y_1, ..., y_m$, which belongs to convex hull of A.

WOHOO!!!

3.2 Problem №2

Show that the set of directions of the strict local descending of the differentiable function in a point is a convex cone.

Solution: It's trivial! Just imagine this picture in your mind, and it will become obvious, in general by definition.

3.3 Problem №3

Prove, that if S is convex, then $S + S = 2S$. Give a counterexample in case, when S – is not convex.

Solution: Let's show that: $2S \subseteq S + S : \forall 2a \in 2S$, $\exists z = x + y \in S + S : z = 2a$, $x, y \in S$. It's true, we can take $x = y = a$. We have proved that $2S \subseteq S + S$

Let's show that: $S + S \subseteq 2S : \forall z \in S + S \exists 2a \in 2S : 2a = z$, $a = 0.5x + 0.5y$, $a \in S$ for reason that S is a convex set. And linear combination of its elements will belong to S. We prove that $S + S \subseteq 2S$.

It means that $S + S = 2S$.

Counterexample: if S isn't convex. $S = \{0\} \bigcup [1, 2]$.
$S + S = [0, 4]$, but $2S = 0 \bigcup [2, 4]$, $S + S \neq 2S$

3.4 Problem №4

Let $x \in \mathbb{R}$ is a random variable with a given probability distribution of $\mathbb{P}(x = a_i) = p_i$, where $i = 1, ..., n$, and $a_1 < ... < a_n$. It is said the probability vector of outcomes of $p \in \mathbb{R}^n$ belongs to the probabilistic simpex, i.e. $P = \{p | \mathbf{1}^T p = 1, p \succcurlyeq 0\} = \{p | p_1 + ... + p_n = 1, p_i \geq 0\}$.
Determine if the following sets of p are convex:

- $\alpha < \mathbb{E}f(x) < \beta$, where $\mathbb{E}f(x)$ stands for expected value of $f(x) : \mathbb{R} \to \mathbb{R}$, i.e. $\mathbb{E}f(x) = \sum\limits_{i=1}^{n} p_i \cdot f(a_i)$
- $\mathbb{E}x^2 \leq \alpha$
- $\mathbb{V}x \leq \alpha$

Solution:

A. It's right because we reduce constraints on p as constraints on the half-space, it will follow from this that the set is convex.

$$\alpha < \mathbb{E}f(x) = \sum_{i=1}^{n} p_i f(a_i) < \beta$$

From the geometry it is half-space and convex, and it means that our set is also convex.

B. Here, we have the same idea like in **A.**. We reduce constraints on p as constraints on the half-space, it will follow this that set is convex.

$$\mathbb{E}x^2 = \sum_{i=1}^{n} p_i a_i^2 \leq \alpha$$

C.

$$0 \leq \mathbb{V}x = \mathbb{E}x^2 - (\mathbb{E}x)^2 = \sum_{i=1}^{n} p_i a_i^2 - \left(\sum_{i=1}^{n} p_i a_i\right)^2 = -p^T X p + d^T p \leq \alpha$$

where $d_i = a_i^2$ and $X = aa^T$, $X \succ 0$. This is a parabola with branches down in multidimensional space, it's concave function. And we also know that it will not be a convex set. For example, $-x^2 < 30$ will not be a convex set.

Answer: **A.-B.** convex. **C** is not convex.

3.5 Problem №5

Let $S \subset \mathbb{R}^n$ is a set of solutions to the quadratic inequality:

$$S = \{x \in \mathbb{R}^n — x^T A x + b^T x + c \leq 0\};\ A \in \mathbb{S}^n, b \in \mathbb{R}^n, c \in \mathbb{R}$$

- Show that if $A \succcurlyeq 0$, S is convex. Is the opposite true?
- Show that intersection of S with the hyperplane defined by the $g^T x + h = 0$, $g \neq 0$ is convex if $A + \lambda g g^T \succcurlyeq 0$ for some $\lambda \in \mathbb{R}$. Is the opposoite true?

Solution:

A.) $x, y \in S$, $\theta \in [0, 1]$. If $\theta = 0$ or 1, it's trivial.

$$\theta x^T A x + \theta b^T x + \theta c \leq 0$$

$$(1-\theta) y^T A y + (1-\theta) b^T y + (1-\theta) c \leq 0$$

$$(\theta x + (1-\theta) y)^T A (\theta x + (1-\theta) y) + b^T (\theta x + (1-\theta) y) + c \leq 0$$

$$\theta^2 x^T A x + \theta(1-\theta)\left(y^T A x + x^T A y\right) + (1-\theta)^2 y^T A y + \theta b^T x + (1-\theta) b^T y + c \leq 0$$

Okay, lets' add $\theta x^T A x + (1-\theta) y^T A y$ and subtract it.
$z = \theta^2 x^T A x + \theta(1-\theta)\left(y^T A x + x^T A y\right) + (1-\theta)^2 y^T A y - \theta x^T A x - (1-\theta) y^T A y$

$$z + \left(\theta x^T A x + \theta b^T x + \theta c\right) + \left((1-\theta) y^T A y + (1-\theta) b^T y + (1-\theta) c\right) \leq 0$$

From that

$$z = \theta^2 x^T Ax + \theta(1-\theta)\left(y^T Ax + x^T Ay\right) + (1-\theta)^2 y^T Ay - \theta x^T Ax - (1-\theta) y^T Ay \leq 0$$

$$\theta(1-\theta)x^T Ax + (1-\theta)\theta y^T Ay \geq \theta(1-\theta)(y^T ax + x^T Ay)$$

$$x^T A(x-y) + (y-x)^T Ay \geq 0$$

It's right because $A \succcurlyeq 0$.

B.) I didn't do it.

4 Convex functions

4.1 Problem №1

Is $f(x) = -x \cdot lnx - (1-x)ln(1-x)$ convex?

Solution:

$$\nabla^2 f(x) = \frac{\partial}{\partial x}\left(-lnx - 1 + ln(1-x) + 1\right) = \frac{-1}{x} - \frac{-1}{1-x} = \frac{-1}{x(1-x)} < 0$$

because $x \in (0,1)$ it is right. From this we get that f(x) is concave function, but not convex.

Answer: No, it's not convex function, it's concave function.

4.2 Problem №2

Let x be a real variable with the values $a_1 < a_2 < ... < a_n$ with probabilities $\mathbb{P}(x = a_i) = p_i$. Derive the convexity or concavity of the following functions from p on the set of $\left\{p — \sum\limits_{i=1}^{n} p_i = 1,\ p_i \geq 0\right)$

Solution: We know that linear function: $a^T x + b$ convex and concave function at the same time.

1. $\mathbb{E}x = \sum\limits_{i=1}^{n} a_i \cdot p_i$ – that's linear function, therefore math expectation is convex and concave function.

2. $\mathbb{P}\{x \geq \alpha\} = \sum\limits_{i: a_i \geq \alpha} p_i$ – it's linear function, therefore it's convex and concave function.

3. $\mathbb{P}\{\alpha \leq x \leq \beta\} = \sum\limits_{i:\alpha \leq a_i \leq b_i}^{n} p_i$, we get that is convex and concave function.

4. $\sum\limits_{i=1}^{n} p_i \log p_i$.

Okay, let's check is $f(x) = x \log x$ - convex

$\nabla^2 f(x) = (x \log x)'' = (\log x + 1)' = \frac{1}{x} > 0$, because $x > 0$ - yep, it's convex function. And our function is non-negative sum of convex function, then our function is convex function.

5. $\mathbb{V}x = \mathbb{E}(x - \mathbb{E}x)^2 = \mathbb{E}x^2 - (\mathbb{E}x)^2$

Okay, let's take counterexample for this function:

1. $p_a = (1, 0)$, $x = (0, 1)$, $\mathbb{V}x = 1 \cdot 1^2 - (1 \cdot 1)^2 = 0$

2. $p_b = (0, 1)$, $x = (0, 1)$, $\mathbb{V}x = 0 \cdot 1^2 + 1 \cdot 0^2 - (1 \cdot 0 + 0 \cdot 1)^2 = 0$

3. $p_c = 0.5 \cdot p_a + 0.5 \cdot p_b = (\frac{1}{2}, \frac{1}{2})$, $x = (0, 1)$, $\mathbb{V}x = 0.5 \cdot 1^2 - (0.5 \cdot 1)^2 = 0.25$

By definition of convex function the following equality is right:

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$

But:

$$f(0.5p_a + 0.5p_b) = \frac{1}{4} \leq \frac{1}{2}f(p_a) + \frac{1}{2}f(p_b) = 0 + 0 = 0$$

And we get that it's not convex function, it's concave function.

6. quatile(x) $= \inf\{\beta — \mathbb{P}\{x \leq \beta\} \geq 0.25\}$

inf is convex function, which is defined on probabilistic simplex, which is convex set, then **quartile**(x) is convex function (Boyd 87 page, statement (3.16)).

Answer: **a-c** convex and concave function, **d.** convex function, **e.** concave function. **f.** convex function.

4.3 Problem №3

Show, that $f(A) = \lambda_{max}(A)$ – is convex, if $A \in S^n$

Solution: Okay, let's show that's is false if $A \in \mathbb{R}^{2n}$:

Let's take:

$$A = \begin{bmatrix} -8 & 16 \\ 60 & 4 \end{bmatrix}, B = \begin{bmatrix} 2 & 40 \\ 20 & -2 \end{bmatrix}$$

$$\lambda_{max}(0.5A+0.5B) = \lambda_{max}\left(\begin{bmatrix} -3 & 28 \\ 40 & 1 \end{bmatrix}\right) \leq 0.5\lambda_{max}\left(\begin{bmatrix} -8 & 16 \\ 60 & 4 \end{bmatrix}\right)+0.5\lambda_{max}\left(\begin{bmatrix} 2 & 40 \\ 20 & -2 \end{bmatrix}\right)$$

$$\lambda_{max}\left(\begin{bmatrix} -3 & 28 \\ 40 & 1 \end{bmatrix}\right), \lambda_{max}\left(\begin{bmatrix} -3 & 28 \\ 40 & 1 \end{bmatrix}\right) = 2(\sqrt{249}-1), \lambda_{max}\left(\begin{bmatrix} -8 & 16 \\ 60 & 4 \end{bmatrix}\right) = 2\sqrt{201}$$

$$\sqrt{1124} - 1 \approx 32.52 \leq \sqrt{249} - 1 + \sqrt{201} \approx 28.96$$

We see that the inequality for a convex function does not hold. Hence, it is not a convex function

Solution: If $A \in \mathbb{S}^n$

We need to consider only diagonal matrix, because all others matrices we can get by multiplying by orthogonal matrices on both sides (they are the same because A - symmetric matrix, $A = \Sigma^T D \Sigma$).

And when we gonna find $\lambda(A) = det(A-\lambda I) = det(\Sigma^T D\Sigma - \lambda \Sigma^T \Sigma) = \lambda(D-\lambda I)$, $D-$ diagonal matrix.

$\forall \theta \in (0, 1)$, for $\theta = 0$ or $\theta = 1$ it's trivial. A, B – diagonal matrices.

$$\theta \cdot A + (1-\theta) \cdot B - \lambda I) = \begin{bmatrix} \theta\lambda_1^a + (1-\theta)\lambda_1^b & 0 & ... \\ 0 & \theta\lambda_2^a + (1-\theta)\lambda_2^b & 0 \end{bmatrix}$$

And from that:

$$\lambda_{max}(\theta A + (1-\theta)B) = \sup_{i=1,n}(\theta\lambda_i^a + (1-\theta)\lambda_i^b) \le \theta \cdot \sup_{i=1,n}(\lambda_i^a) + (1-\theta)\sup_{i=1,n}(\lambda_i^b)$$

$$\theta \cdot \sup_{i=1,n}(\lambda_i^a) + (1-\theta)\sup_{i=1,n}(\lambda_i^b) = \theta\lambda_{max}(A) + (1-\theta)\lambda_{max}(B)$$

Wohoo, we proved that λ_{max} is convex function.

Alternative solution: $\lambda_{max}(A) = \max\limits_{||u||=1}\langle u, Au\rangle$.

We know that $f(A) = \langle u, Au\rangle$ is linear function, which means it's convex function. max is convex non-decreasing function. And from that we get that composition of this function is convex function.

Wohoo, we proved that λ_{max} is convex function.

4.4 Problem №4

Prove that, $f(X) = -\log\det X$ is convex on $X \in \mathbb{S}^n_{++}$

Solution:

$$df(x) = -\frac{1}{detX}detX\langle X^{-T}, dX\rangle = -\langle X^{-T}, dX\rangle$$

$$d^2f(x) = -d(tr(X^{-1}, dX_1) = -tr(d(X^{-1})dX_1) = tr(X^{-1}dX_2X^{-1}dX_1)$$

For a reason that $X^{-1}, dX_1, dX_2 \in \mathbb{S}^n_{++}$ trace is positive, the hessian of f(x) is positive, then we get that f(x) is convex function on $\mathbb{S}^n_{++}$. In later I understand that dX_1 and dX_2 can be non positive matrices. And this is an unfair statement.

Alternative solution: Let's take: $X = A + tB$, $A \in \mathbb{S}^n_{++}, B \in \mathbb{S}^n$, t is such for which $A + tB \succ 0$. We can represent A as $A = CC$, $det(A) = det(CC) = (detC)^2$. $F(t) = \log det(A + tB)$.

$$F(t) = \log det(A + tB) = \log det(CC + tB) = \log det(C(I + tC^{-1}BC^{-1})C)$$

$$F(t) = \log det(I + tC^{-1}BC^{-1}) + \log detA$$

Let's write $C^{-1}BC^{-1} = U\Sigma V^*$, $\Sigma = diag(\lambda_1, ..., \lambda_n)$, because it's symmetric, $U = V$, then:

$$F(t) = \log det(UU^* + tU\Sigma U^*) + \log detA = \log det(U(I + t\Sigma)U^*) + \log detA$$

$$F(t) = \log det(I+t\Sigma) + \log detA = \log\prod_{i=1}^{n}(1+t\lambda_i) + \log detA = \sum_{i=1}^{n}\log(1+t\lambda_i) + \log detA$$

Then F(t) is concave function, because it's sum of concave functions. Then we get that $-F(t)$ is convex function. Wohoo, we proved it.

4.5 Problem №5

Prove, that adding $\lambda||x||_2^2$ to any convex function $g(x)$ ensures strong convexity function of a resulting function $f(x) = g(x) + \lambda||x||_2^2$. Find the constant of the strong convexity μ.

Solution:

$$\frac{\partial^2}{\partial x_l \partial x_k} \sum_{i=1}^{n} x_i = \frac{\partial}{\partial x_l} 2x_k = 0$$

$$\frac{\partial^2}{(\partial x_k)} \sum_{i=1}^{n} x_i = \frac{\partial}{\partial x_k} 2x_k = 2$$

The hessian of $||X||_2^2$ is $2I$. Then $\nabla^2 f(x) = \nabla^2 g(x) + \lambda \cdot 2I \succcurlyeq \lambda I$. It's right because g(x) is convex function.

4.6 Problem №6

Study the following function of two variables $f(x, y) = e^{xy}$.

a. Is this function convex?

b. Prove, that this function will be convex on the line $x = y$.

c. Find another set in $\mathbb{R}^2, on which this function will be convex$

Solution:

a. No this function is not convex on $\mathbb{R}^2$, because the hessian equals $-(1+2xy)e^{2xy}$ and it is less than zero for $x = 0, y = 1$.

b. $f(x, x) = e^{x^2}$, $\nabla^2 f(x) = 2(1 + 2x^2)e^{x^2} > 0$, $\forall x \in \mathbb{R}$, and from that we get that $f(x)$ is strong convex function.

c. Let's take $y = x^3$, $f(x, x) = e^{x^4}$, $\nabla^2 f(x) = (12x^2 + 16x^6)e^{x^4} \geq 0$, $\forall x \in \mathbb{R}$

Answer: **a.** No, this function is not convex on $\mathbb{R}^2$, **b.** proved, **c.** $y = x^3$

4.7 Problem №7

Show, that the following function is convex on the set of all positive denominators:

$$f(x) \frac{1}{x_1 - \frac{1}{x_2 - \frac{1}{x_3 - \frac{1}{\dots}}}}, \ , x \in \mathbb{R}^n$$

Solution: $f_n = \frac{1}{x_n}, f_{n-1} = \frac{1}{x_{n-1}-f_n}$

Let's make induction from n to n-1: $(f_n)'' = \frac{2}{x_n^3} > 0$, $x_n > 0$.

$n-1$: $f_{n-1} = \frac{1}{x_{n-1}-f_n}$,

$$\nabla^2 f_{n-1}(x) = \begin{bmatrix} \frac{2}{(x_{n-1}-f_n)^3} & \frac{2}{x_n^2(x_{n-1}-f_n)^3} \\ \frac{2}{x_n^2(x_{n-1}-f_n)^3} & \frac{2}{x_n^3(x_n-f_n)^2} + \frac{2}{x_n^4(x_{n-1}-f_n)^3} \end{bmatrix}$$

1. $\frac{2}{(x_{n-1}-f_n)^3} > 0$ it's true by a condition of a task.
2. $|\nabla^2 f(x)| = \frac{4}{x_n^3(x_{n-1}-f_n)^3} + \frac{4}{x_n^4(x_{n-1}-f_n)^6} - \frac{4}{x_n^4(x_{n-1}-f_n)^6} = \frac{4}{x_n^3(x_{n-1}-f_n)^3} > 0$

Wohoo, $\nabla^2 f_{n-1}(x)$ is positive defined matrix, it's means that f_{n-1} is convex function.

By induction we get that $f_1(x)$ is convex function.

Alternative solution: Exists a statement: if $\phi(x)$ is convex function that is monotonously decreasing and $\alpha(x)$ is concave function, than $f(x) = \phi(\alpha(x))$ will be convex function. For our example: $\phi(x) = 1/x)$, $\phi(x)$ – convex function, and $\alpha_k(x) = x_k - \phi(a_{k+1})$ is concave function, $\phi(x) = \alpha_n(x) = 1/x$, $-\alpha_i(x)$ is concave function.

We can represent f in the form of: $f(x) = \phi(x_1 - \phi(x_2 - \phi(x_3 - ...\phi(x_{n-1} - \phi(x_n))...)$ will be convex function based on this statement. (3.2.4 Scalar composition, 84 page Optimization Boyd)

5 Conjugate sets

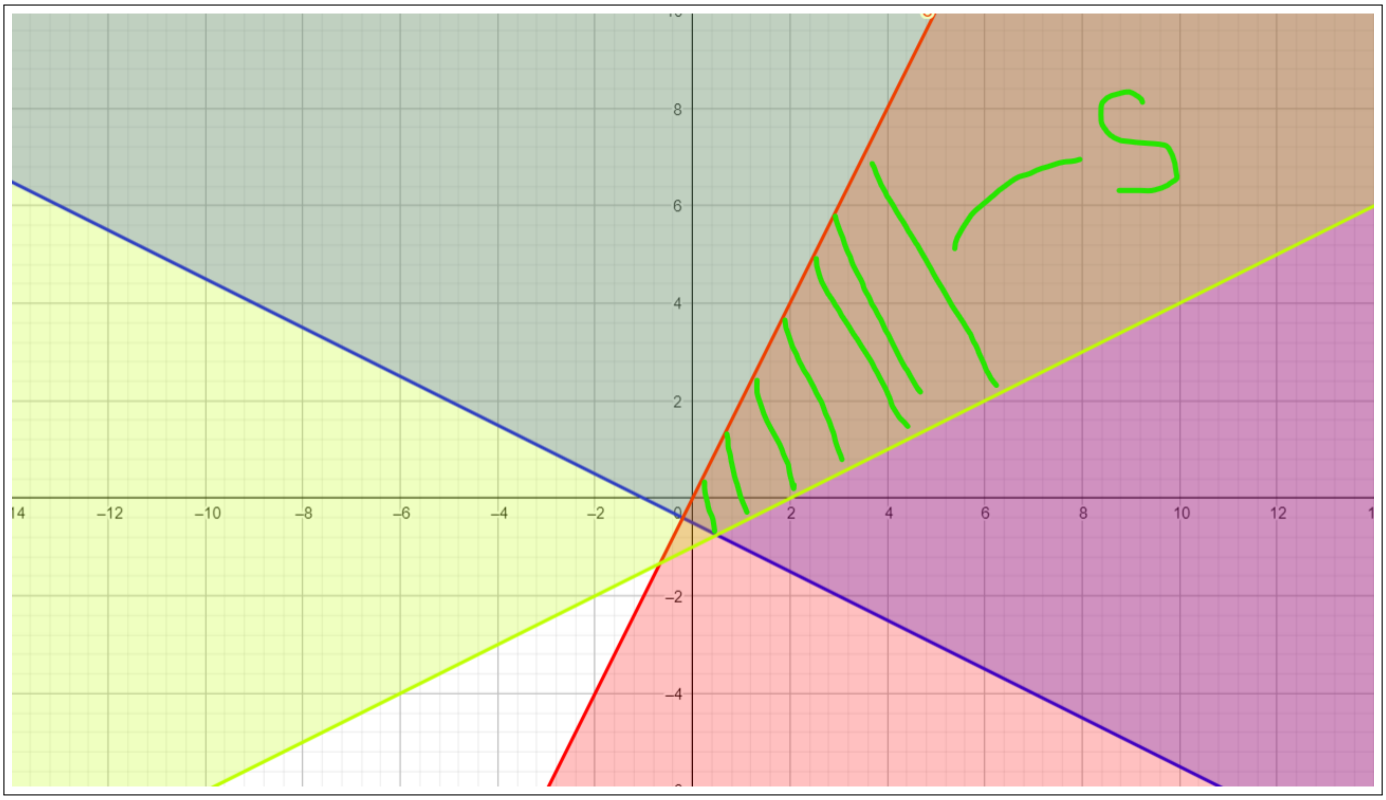
5.1 Problem №1

Find the sets $\mathbb{S}^*, \mathbb{S}^{**}, \mathbb{S}^{***}$, if

$$\mathbb{S} = \{x \in \mathbb{R}^2 — x_1 + x_2 \geq -1,\ 2x_1 - x_2 \geq 0,\ \ -x_1 + 2x_2 \leq -2\}$$

Solution: S is closed, convex and includes 0, then $S^{**} = S$, $S^{***} = S^*$.

$$S = conv\left\{\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} \frac{-1}{3} \\ \frac{-2}{3} \end{pmatrix}\right\} + cone\left\{\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right\}$$



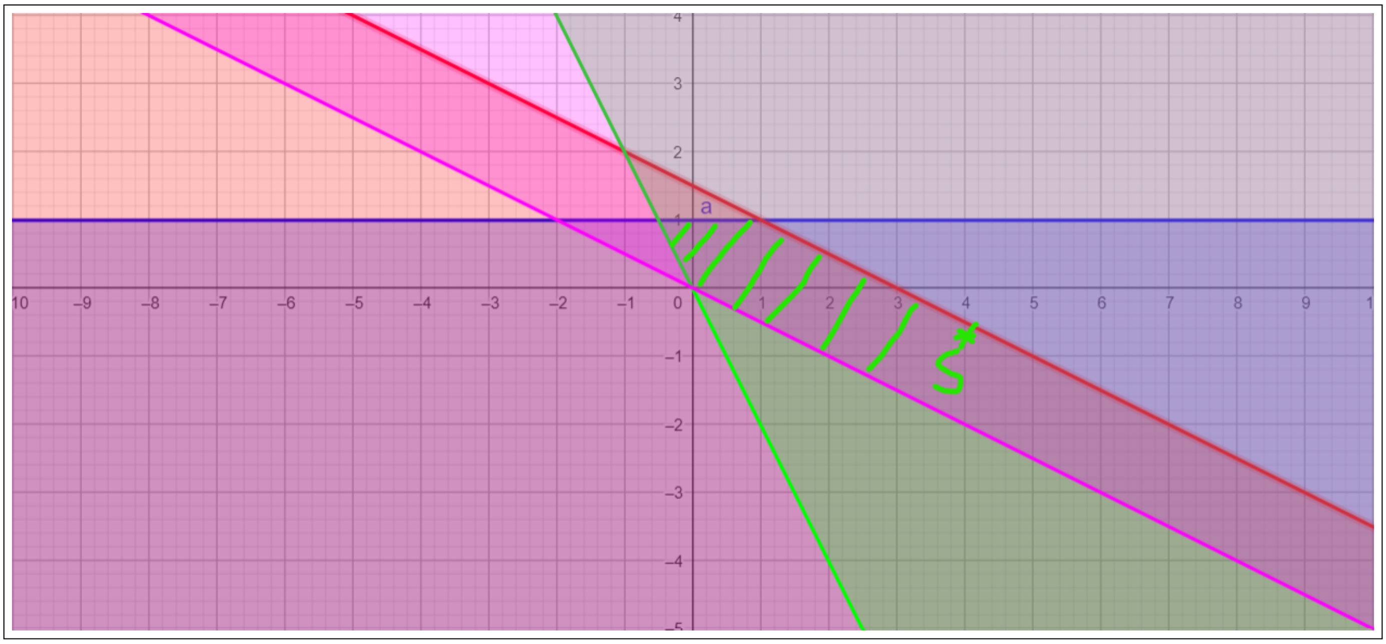
And then we get this:

$$\begin{cases} 0 \cdot p_1 - p_2 \geq -1 \\ -\frac{1}{3}p_1 - \frac{2}{3}p_2 \geq -1 \\ 2p_1 + p_2 \geq 0 \\ p_1 + 2p_2 \geq 0 \end{cases}$$

$$\begin{cases} p_2 \leq 1 \\ p_1 + 2p_2 \leq 3 \\ 2p_1 + p_2 \geq 0 \\ p_1 + 2p_2 \geq 0 \end{cases}$$

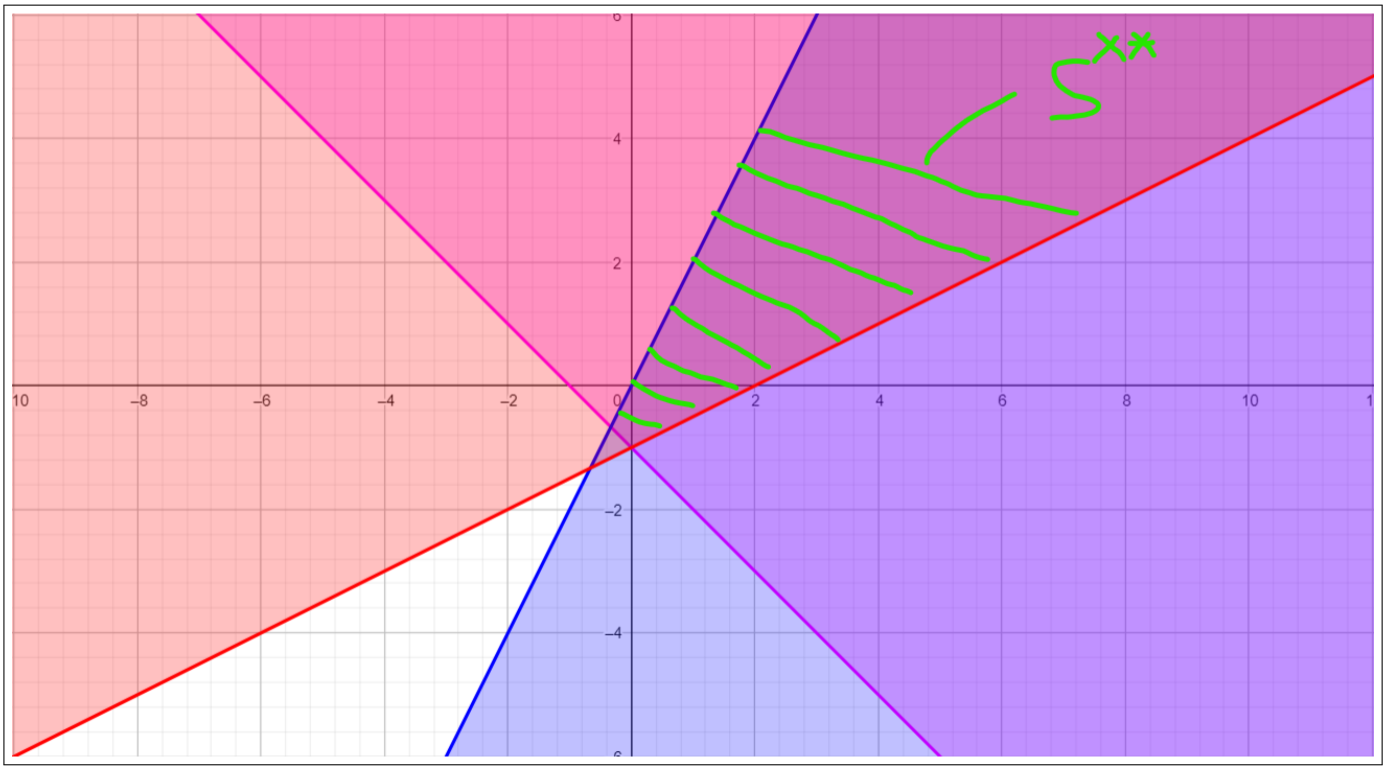
And we get:

$$S^* = conv\left\{\begin{pmatrix}\frac{-1}{2}\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}\right\} + cone\left\{\begin{pmatrix}2\\-1\end{pmatrix}\right\}$$



$$\begin{cases}-\frac{1}{2}p_1 + p_2 \geq -1\\ p_1 + p_2 \geq -1\\ 2p_1 - p_2 \geq 0\end{cases}$$

$$S^{**} = conv\left\{\begin{pmatrix}0\\-1\end{pmatrix}, \begin{pmatrix}\frac{-1}{3}\\\frac{-2}{3}\end{pmatrix}\right\} + cone\left\{\begin{pmatrix}2\\1\end{pmatrix}, \begin{pmatrix}1\\2\end{pmatrix}\right\}$$



S^{***} the same as S^*. I draw S^{**} to demonsrate that they are similar.

Answer: All answers are above.

5.2 Problem №2

Prove, that K_p and K_{p^*} are inter-conjugate, i.e. $(K_p)^* = K_{p^*}$, $(K_{p^*})^* = K_p$, where $K_p = \left\{[x, \mu] \in \mathbb{R}^{n+1} : ||x||_p \leq \mu\right\}$, $1 < p < \infty$ is the norm cone (w.r.t p - norm) and p, p_* are conjugated, i.e. $p^{-1} + p_*^{-1} = 1$. You can assume, that $p_* = \infty$ if $p = 1$ and viceversa.

Solution: $K := \{x \in \mathbb{R}^n \mid ||x|| \leq 1\}$ Unit ball is convex, symmetric by x and closed, norm we can write in such way: $||x|| = \frac{1}{\sup\{t \geq 0 \mid tx \in K\}}$.

From properties of conjugate norm: $(||x||_p)^* = ||x||_{p^*}$, if $p^{-1} + p_*^{-1} = 1$.

We have finite-dimensional space, then will be executed $||x||_{**} = ||x||, \forall x$

Wohoo, we proved.

5.3 Problem №3

The cone $\mathbb{R}^n_+$ fits for it. $x^T y \geq 0 \ \forall x \succcurlyeq 0 \iff y \succcurlyeq 0$ and it's $\mathbb{R}^n_+$.

5.4 Problem №4

Find the conjugate set to the ellipsoid:

$$S = \left\{ x \in \mathbb{R}^n | \sum_{i=1}^n a_i^2 x_i^2 \leq \varepsilon^2 \right\}$$

Solution: It's equivalent to:

$$S = \left\{ x \in \mathbb{R}^n | \sum_{i=1}^n a_i^2 x_i^2 \leq \varepsilon^2 \right\}$$

$A = diag(\frac{a_i}{\varepsilon})$

$$||Ax||_2^2 = \sum_{i=1}^n (\frac{a_i}{\varepsilon})^2 x_i^2$$

From Boyd's optimization I know that:

$$S = \{x \in \mathbb{R}^n \mid ||Ax||_2 \leq 1\} = E = \{A^{-1}u \mid ||u||_2 \leq 1\}$$

$A^{-1} = diag(\frac{\epsilon}{a_i})$.

Let's show that it's true.

1. $E \subseteq S$: $x = A^{-1}u$, $||u||_2 \leq 1$ in S: $||AA^{-1}u||_2 \leq 1$, therefore $E \subseteq S$ it's true.
2. $S \subseteq E$: $||Ax||_2 \leq 1$, $z = Ax$, $||z||_2 \leq 1$, $x = A^{-1}z$, $||z||_2 \leq 1$, therefore $x \ \in E$, $\hookrightarrow S \subseteq E$

We prove that it's true.
We need find all $p \in E^* : \forall x \in E,\ \langle p, x\rangle \geq -1,\ \forall x = A^{-1}u,\ ||u||_2 \leq 1,$

$$\langle p, A^{-1}u\rangle = \langle A^{-T}p, u\rangle \geq -1$$

From cauchy-bunyakovsky and $||u||_2 \leq 1$ we get:

$$|\langle A^{-T}p, u\rangle| \leq ||u||_2 \cdot ||A^{-T}p||_2 \leq ||A^{-T}p||_2$$

We are suitable for all p, for which is true: $\langle A^{-T}p, u\rangle = -||A^{-T}p||_2$, $-||A^{-T}p||_2 \geq -1$, we get:

$$||A^{-T}p||_2 \leq 1$$

This sets an ellips with matrix $A^{-T} = diag(\frac{\varepsilon}{a_i})$.

$$S^* = E^* = \left\{ x \in \mathbb{R}^n | \sum_{i=1}^{n} (\frac{\varepsilon}{a_i})^2 x_i^2 \leq 1 \right\}$$

Answer:

$$S^* = \left\{ x \in \mathbb{R}^n | \sum_{i=1}^{n} (\frac{1}{a_i})^2 x_i^2 \leq \frac{1}{\varepsilon^2} \right\}$$

5.5 Problem №5

Find the conjugate cone for the exponential cone:

$$K = \left\{ (x, y, z) | y > 0,\ ye^{\frac{x}{y}} \leq z \right\}$$

Solution: By definition of conjugate set:

$$\begin{cases} ax + by + cz \geq 0 \\ y > 0 \\ ye^{\frac{x}{y}} \leq z \end{cases} \text{, let's take: } t = \frac{x}{y}, p = \frac{z}{y}, \begin{cases} at + b + cp \geq 0 \\ p \geq e^t \\ p \leq 0 \end{cases}$$

1. $c = 1$, let's take a and b in such way that all values lie above $p \geq e^t$.
   (a) $a > 0$: $\nexists\, b$
   (b) $a = 0$: $b \geq 0$
   (c) $a < 0$: $\begin{cases} p \geq e^t \\ p \geq -b - at \end{cases}$, $\begin{cases} p' = e^t \\ p' = -a \end{cases}$, $\begin{cases} t = \ln(-a) \\ p = -a \end{cases}$, $b \geq a(1 - \ln(-a))$
2. $c = 0$, $\nexists\, a, b$ such that all values of $at + b \geq 0$ lie above $p \geq e^t$. It's vertical line.

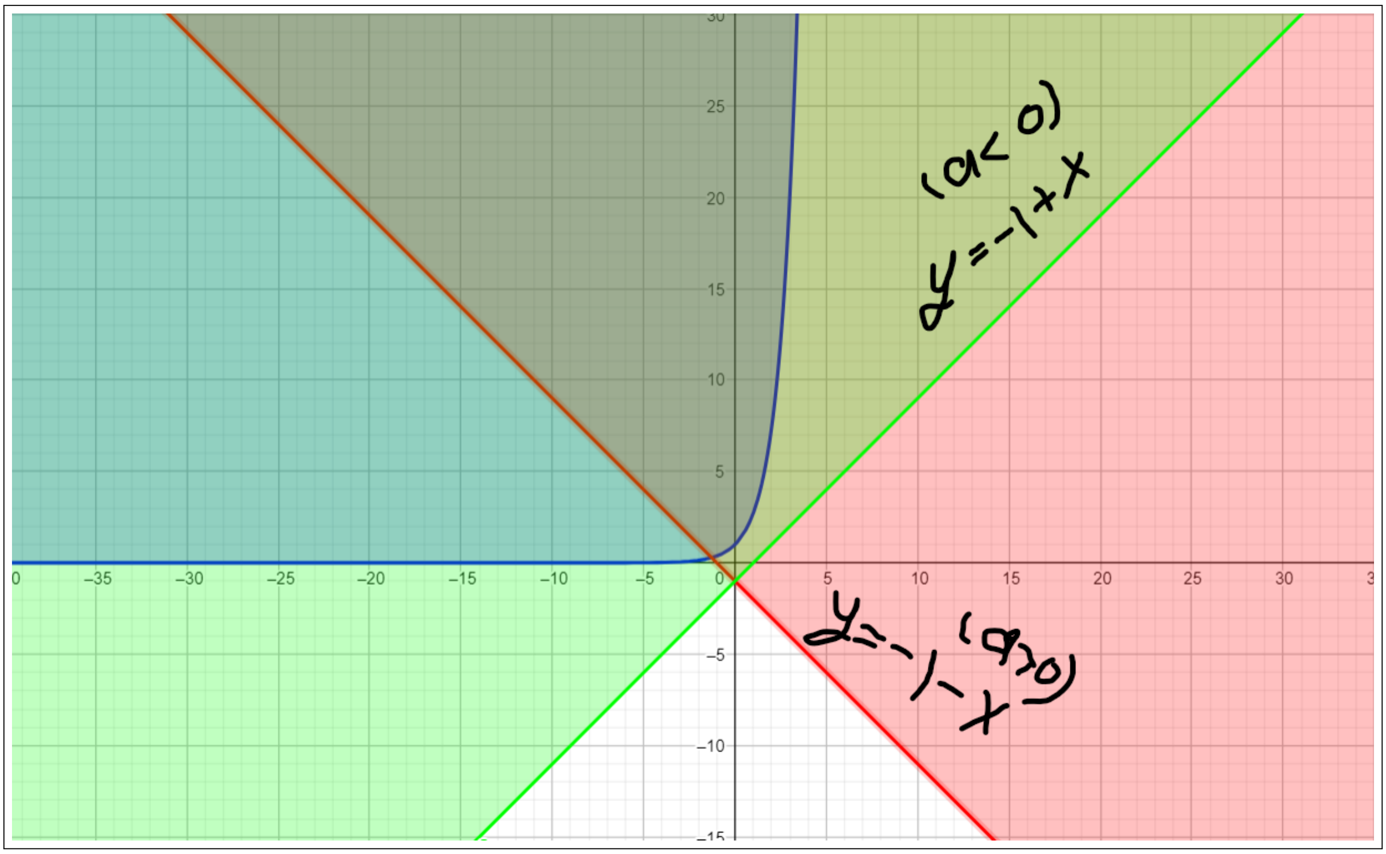


Figure 1: $c = 1$

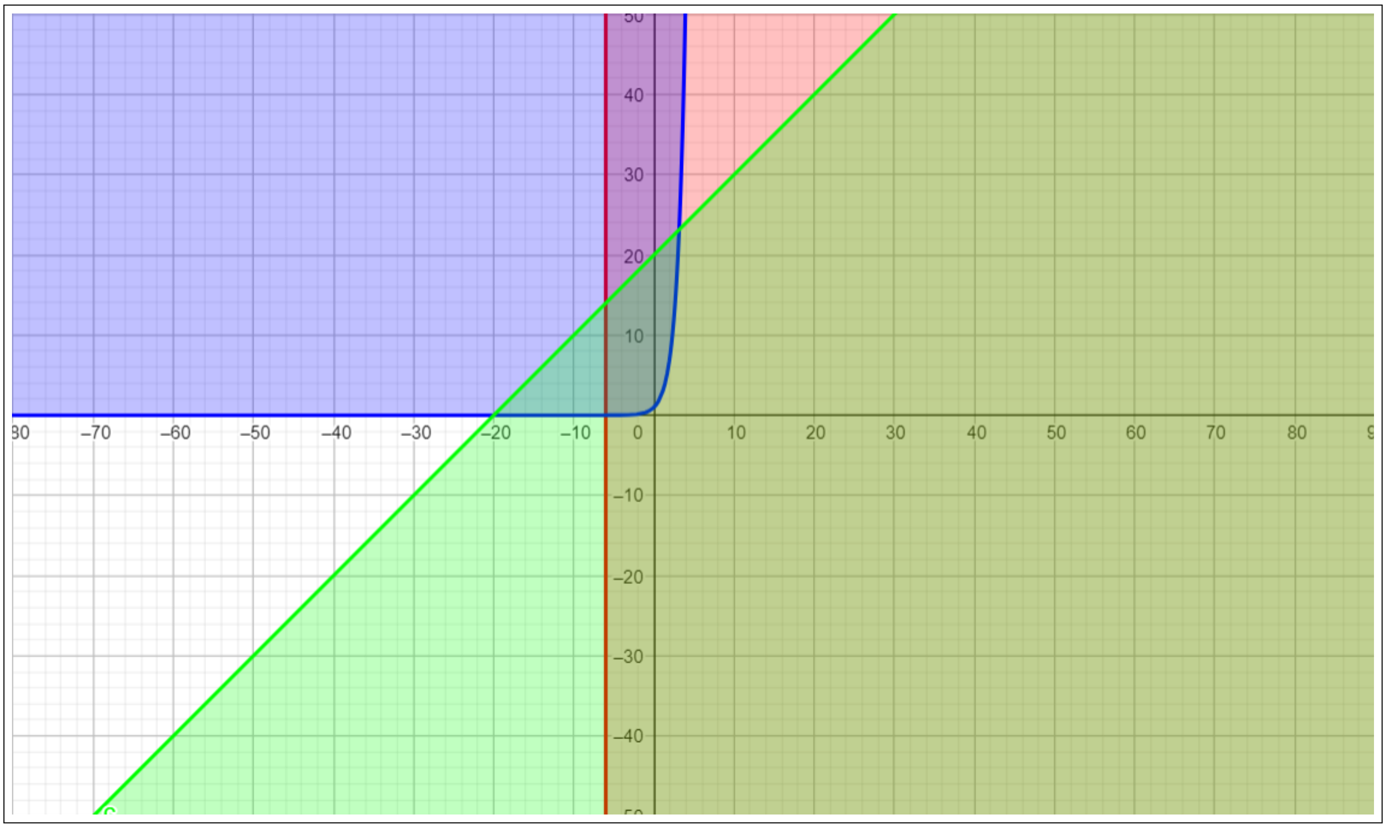


Figure 2: $c = 0$ and $c < 0$

3. $c = -1$, All points $p \geq e^t$ must lie above $p \leq at + b$, but $\nexists\, a, b$.

   All others solutions could get by mupltipluyng c on some positive constant.
   Answer:

$$K^* = \{(c \cdot a, c \cdot b, c) — c \geq 0, \text{ if } a = 0 \text{ and } b \geq 0, \text{ or } a < 0 \text{ and } b \geq a(1 - \ln(-a))\}$$

6 Conjugate functions

6.1 Problem №1

Find $f^*(y)$, if $f(x) = p \cdot x - q$

Solution:

$$f^*(y) = \sup_{x \in dom(f)} (\langle y, x\rangle - f(x)) = \sup_{x \in \mathbf{R}} (xy - px + q)$$

$g(x) = xy - px + q$, $\nabla g(x) = y - p$ And we can easy get a answer:

Answer: $f^*(y) = \begin{cases} q, y = p \\ +\inf, y \neq p. \end{cases}$

6.2 Problem №2

Find $f^*(y)$, if $f(x) = \frac{1}{2}x^T Ax$, $A \in \mathbf{S}^n_{++}$

Solution:

$$f^*(y) = \sup_{x \in dom(f)} (\langle y, x\rangle - f(x)) = \sup_{x \in \mathbf{R}^n} \left(y^T x - \frac{1}{2}x^T Ax\right)$$

$g(x) = y^T x - \frac{1}{2}x^T Ax$

$\nabla g(x) = y - Ax$, because $A \in \mathbf{S}^n_{++}$, $\nabla g(x) = 0 \hookrightarrow y = Ax$, $x = A^{-1}y$

$$f^*(y) = y^T A^{-1}y - \frac{1}{2}(y^T A^{-1}AA^{-1}y = \frac{1}{2}y^T A^{-1}y$$

Answer: $f^*(y) = \frac{1}{2}y^T A^{-1}y$

6.3 Problem №3

Find $f^*(y), iff(x) = log\left(\sum_{i=1}^n e^{x_i}\right)$

Solution:

$$f^*(y) = \sup_{x \in dom(f)} (\langle y, x\rangle - f(x)) = \sup_{x \in \mathbf{R}^n} \left(y^T x - log\left(\sum_{i=1}^n e^{x_i}\right)\right)$$

$d(y^t x - f(x)) = \langle y, dx\rangle - \frac{\langle e^x, dx\rangle}{\sum_{i=1}^n e^{x_i}} = 0$, e^x in meaning that it equals $(e^{x_1}, e^{x_2}, ..., e^{x_n})$.

And we get that: $y = \frac{e^x}{\sum_{i=1}^n e^{x_i}}$

$$f^*(y) = \sup_{x \in \mathbf{R}^n} \left(\sum_{i=1}^n y_i x_i - log\left(\sum_{i=1}^n e^{x_i}\right)\right) = \sup_{x \in \mathbf{R}^n} \left(\sum_{i=1}^n log(e^{y_i x_i}) - log\left(\sum_{i=1}^n e^{x_i}\right)\right)$$

$$f^*(y) = \sup_{x\in\mathbf{R}^n}\left(\sum_{i=1}^{n} log\left(\frac{e^{y_i}e^{x_i}}{\sum\limits_{k=1}^{n} e^{x_k}}\right)\right) = \sum_{i=1}^{n} log\left(e^{y_i}y_i\right) = \sum_{i=1}^{n} y_i log(y_i)$$

Oh wow, it's crossentropyloss ($y \succ 0$, $y^T\mathbb{1} = 1$, so y - probability vector)
Now we need to find dom f^*.

1. if $\exists y_i : y_i < 0$, let's take $x_j = -\alpha$, and $x_l = 0$ for all l : $l \neq j$, then:

$$y^Tx - f(x) = \alpha - log\alpha \xrightarrow{\alpha\to+\infty} +\infty$$

2. if $y \succ 0$, but $y^T\mathbb{1} \neq 1$, let's take $x = \alpha \cdot \mathbb{1}$

$$y^Tx - f(x) = y^T\alpha\mathbb{1} - \alpha - \log(n)$$

if $y^T\mathbb{1} > 1$, then we take $\alpha \to +\infty$ and get that $f^*(y) = +\infty$
if $y^T\mathbb{1} < 1$, then we take $\alpha \to -\infty$ and get that $f^*(y) = +\infty$

Answer:

$$f^*(y) = \begin{cases} \sum\limits_{i=1}^{n} y_i \cdot log(y_i), \text{if } y \succ 0 \text{ and } y^T\mathbb{1} = 1 \\ +\infty, otherwise \end{cases}$$

6.4 Problem №4

Prove, that if $f(x) = g(Ax)$, then $f^*(y) = g^*(A^{-T}y)$
Solution:

$$f^*(y) = \sup_{x\in\mathbf{R}^n}\left(y^Tx - f(x)\right)$$

$y^Tdx - \nabla f(x)^Tdx = 0$, $y = \nabla f(x)$

$$f^*(y) \sup_{x\in\mathbf{R}^n}\left((\nabla f(x))^Tx - f(x)\right)$$

The same way we can get:

$$g^*(y) = \sup_{x\in\mathbf{R}^n}\left((\nabla g(x))^Tx - g(x)\right)$$

$$g^*(A^{-T}y) = \sup_{x\in\mathbf{R}^n}\left((A^{-T}\nabla(g(x))^Tx - g(x)\right)$$

$x = Ax_1$

$$g^*(A^{-T}y) = \sup_{x_1 \in \mathbf{R}^n} \left(\nabla g(Ax_1)^T A^{-1} A x_1 - g(Ax_1)\right) = \sup_{x_1 \in \mathbf{R}^n} \left(\nabla g(Ax_1)^T x_1 - g(Ax_1)\right)$$

Because $g(Ax_1) = f(x_1)$, we can get:

$$g^*(A^{-T}y) = \sup_{x_1 \in \mathbf{R}^n} \left(\nabla f(x_1)^T x_1 - f(x_1)\right) = f^*(y)$$

WOHOO!

6.5 Problem №5

Find $f^*(Y)$, if $f(X) = -\log(detX), X \in \mathbf{S}^n_{++}$

Solution:

$$f^*(Y) = \sup_{x \in \mathbf{R}^{n \times n}} (\langle Y, X\rangle + \log(detX))$$

$\langle Y, dX\rangle + \frac{1}{detX} d(detX) = \langle Y, dX\rangle + \frac{1}{detX} detX \langle X^{-T}, dX\rangle = 0$
And we get: $Y = -X^{-T}$, $X = -Y^{-T}$

$$f^*(Y) = \langle Y, -Y^{-T}\rangle + \log(det(-Y^{-1})) = tr(-E) + \log\left(det(-Y^{-1})\right)$$

$$f^*(Y) = -n + \log\left(det(-Y^{-1})\right)$$

Answer:

$$f^*(Y) = -n + log(det(-Y^{-1})), \text{ where } Y \in -\mathbb{S}^n_{++}$$

7 Subgradient and subdifferential

7.1 Problem № 1

Find $\partial f(x)$, if $f(x) = \textbf{Leaky ReLU}(x) = \begin{cases} x, if x > 0 \\ 0.01x, otherwise \end{cases}$

Solution: By Dubovitsky-Milutin theorem we can get:

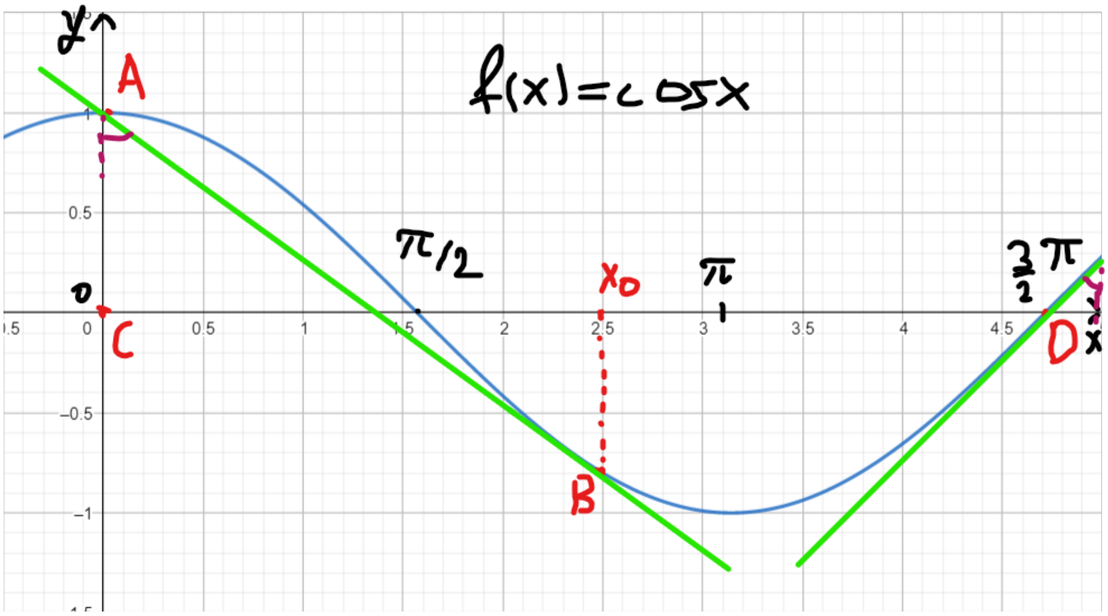
$$\partial f(x) = \begin{cases} 1, x > 0 \\ [0.01; 1], x = 0 \\ 0.01, x < 0 \end{cases}$$

Answer: $\partial f(x) = \begin{cases} 1, x > 0 \\ [0.01; 1], x = 0 \\ 0.01, x < 0 \end{cases}$

7.2 Problem №2

Find subdifferential of a function $f(x) = cosx$ on the set $X = [0, \frac{3}{2}\pi]$.

Solution:



Answer: $\partial f(x) = \begin{cases} [-\infty, -sinx], x = 0 \\ \varnothing, x \in (0, x_0) \\ -sinx, x \in [x_0, \frac{3}{2}\pi) \\ [1, +\infty], x = \frac{3}{2}\pi \end{cases}$

7.3 Problem №3

Find $\partial f(x)$, if $f(x) = ||Ax - b||_1^2$

Solution: By property of subdifferential we can get that:

$$\partial(||Ax - b||_1^2) = ||Ax - b||_1 \partial(||Ax - b||_1)$$

From the seminar we know that:

$$\partial||y||_1 = \{\alpha : ||\alpha||_\infty \leq 1, \alpha^T y = ||y||_1\}$$

And now we can get that (by another property of subdifferential:

$$\partial\left(||Ax - b||_1^2\right) = ||Ax - b||_1 \partial\left(||Ax - b||_1\right)(x) = ||Ax - b||_1 A^T \partial||Ax + b||_1$$

$$||Ax - b||_1 A^T \partial||Ax + b||_1 = ||Ax - b||_1 A^T \cdot \{\alpha : ||\alpha||_\infty \leq 1, \alpha^T(Ax + b) = ||Ax + b||_1\}$$

Answer: $\partial f(x) = ||Ax - b||_1 A^T \cdot \{\alpha : ||\alpha||_\infty \leq 1,\ \alpha^T(Ax + b) = ||Ax + b||_1\}$

7.4 Problem №4

Suppose, that if $f(x) = ||x||_\infty$. Prove that $\partial f(0) = \textbf{conv}\ \{\pm e_1, ..., \pm e_n\}$, where e_i is i-th canonical basis vector (column of identity matrix).

Solution: By the definition: $f(x) = ||x||_\infty = \max\limits_i |x_i|$

We know that subbdifferential for module is equal:

$$\partial|x_i| = \begin{cases} x_i, x_i > 0 \\ [-1, 1], x_i = 0 \\ -x_i, x_i < 0 \end{cases}$$

Because $|x_i|$ – convex functions, by Dubovitsky - Milutin theorem we can get:

$$\partial f(0) = conv\left\{\bigcup_{i \in \overline{1,n}} \partial|x_i|_{x_i=0}\right\} = conv\left\{\pm e_1, ..., \pm e_n\right\},\ e_i \text{ is i-th canonical basis vector}$$

WOHOO, we proved that!

7.5 Problem №5

Find $\partial f(x)$, if $f(x) = e^{||x||}$.

Try do the task for an arbitrary norm. At least, try $||\cdot|| = ||\cdot||_{\{2,1,\infty\}}$

Solution:

By the property of subdifferential we: $\partial f(x) = \partial(e^{||x||}) = e^{||x||}\partial(||x||)$

And now we need to find $\partial||x||$ for $||x||_1, ||x||_2$ and $||x||_\infty$
1. In the seminar we find that and it equals:

$$\partial||x||_1 = \{\alpha : ||\alpha||_\infty \leq 1, \alpha^T x = ||x||_1\}$$

2. By definition $||x||_2 = \sqrt{\sum\limits_{i=1}^{n} x_i^2}$, this function is differentiable everywhere except zero.
For $x \neq 0$, $\partial f(x) = \nabla||x||_2 = \frac{x}{||x||_2}$
Now we need to consider $x = 0$: let's find such interesting limit: $\lim\limits_{\beta \longrightarrow 0+} \frac{||\beta e||_2}{\beta} = ||e||_2$, where e is unit vector on unit sphere.
But by definition $\frac{x}{||x||_2} = e$, and we get that

$$\partial||x||_2 = \{e — ||e||_2 \leq 1\}$$

3. In problem №4 I find that (x_i - maximum element by modules):

$$||x||_\infty = \begin{cases} [-1, 1], \text{ if } ix_i = 0 \\ sign(x_i), x_i \neq 0 \end{cases}$$

Answer:

- for $||\cdot||_1$: $\partial f(x) = e^{||x||_1} \cdot \{\alpha : ||\alpha||_\infty \leq 1, \alpha^T x = ||x||_1\}$
- for $||\cdot||_2$: $\partial f(x) = e^{||x||_2} \cdot \{e — ||e||_2 \leq 1\}$
- for $||\cdot||_\infty$: $\partial f(x) = e^{||x||_\infty} \cdot \begin{cases} [-1, 1], \text{ if } x_i = 0 \\ sign(x_i), x_i \neq 0 \end{cases}$, where is x_i – maximum element by modules

8 General optimization problem

8.1 Problem №1

Consider the problem of projection some point $y \in \mathbb{R}^n, y \notin \Delta^n$ onto the unit simplex Δ^n. Find 2 ways to solve the problem numerically and compare them in terms of the total computational time, memory requirements and iteration number for $n = 10, 100, 1000$

$$||x - y||_2^2 \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } \mathbb{1}^T x = 1, x \succeq 0$$

Solution: there is no analytical solution. I had no time for do it. But I think that problem is $O(n^3)$.

8.2 Problem №2

Give an explicit solution of the following LP.

$$c^T x \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } Ax = b$$

Solution:
Secondly, we can write: $L(x, \nu) = c^T x + \nu(Ax - b)$, then write KKT conditions:

$$\begin{cases} \nabla_x L = c + A^T \lambda = 0 \\ \nabla_\nu L = Ax - b = 0 \\ \nu \succeq 0 \end{cases}$$

It's convex optimization problem all functions are affine.

1. Firstly, if there is no solution to $Ax = b$, then budget set is empty and $p^* = +\infty$.
2. Secondly, if there is solution to $Ax = b$, then $x^* = A^\dagger b$ and $p^* = c^T A^\dagger b$

Answer:

$$\begin{cases} c^T A^\dagger b, \text{ if there is no solution for } Ax = b \\ +\infty, \text{ otherwise} \end{cases}$$

8.3 Problem №3

Give an explicit solution of the following LP.

$$c^T x \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } \mathbb{1}^T x = 1, x \succeq 0$$

This problem can be considered as a simplest portfolio optimization problem.

Solution: Lagrangian:

$$L(x, \nu, \lambda) = c^T x + \nu(\mathbb{1}^T x - 1) - \lambda^T x$$

Let's write KKT conditions:

$$\begin{cases} \nabla_x L = c + \nu \mathbb{1} - \lambda = 0 \\ \nabla_\nu L = \mathbb{1}^T x - 1 = 0 \\ \lambda \succeq 0 \\ \lambda_k x_k = 0, k \in \overline{1, n} \\ x \succeq 0 \end{cases}$$

Okay, we can find minimal component of c (c_k). It's clear that such component exists. And if we take $(\mathbb{1}^T x) c_k \leq c^T x$, for all x which satisfy for conditions.

And we easily get: $p^* = c_k$. It's funny that we need to invest all our money in only one asset, but it's clearly true if we know and sure in all outcomes. Such λ and ν exist that satisfy KKT, $\nu^* = -c_k$, $\lambda_i^* = c_i - c_k, i \in \overline{1, n}$

Answer: $p^* = c_k$, $x^* = (0, ..., 1, ...0)$

8.4 Problem №4

Give an explicit solution of the following LP.

$$c^T x \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } \mathbb{1}^T x = \alpha, 0 \preccurlyeq x \preccurlyeq 1$$

where α is an integer between 0 and n. What happens if α is not an integer (but satisfies $0 \leq \alpha \leq n$)? What if we change the equality to an inequality $\mathbb{1}^T x \leq \alpha$?

Solution: Lagrangian:

$$L(x, \nu, \lambda_1, \lambda_2) = c^T x + \nu(\mathbb{1}^T x - \alpha) + \lambda_1^T (x - \mathbb{1}) + \lambda_2^T (-x)$$

KKT conditions:

$$\begin{cases} \nabla_x L = c + \nu \mathbb{1} + \lambda_1 - \lambda_2 = 0 \\ \nabla_\nu L = \mathbb{1}^T x - 1 = 0 \\ \lambda_1 \succeq 0 \\ \lambda_2 \succeq 0 \\ \lambda_{1_k}(x_k - 1) = 0, k \in \overline{1, n} \\ \lambda_{2_k} x_k = 0, k \in \overline{1, n} \\ 1 \succeq x \succeq 0 \end{cases}$$

KKT conditions will be necessary and sufficient because Slater's condition is met.

If α is integer, then we consider without begging for generality without begging for generality (from previous task) $k = 0$, and the smallest components come first.

We take first α components of c: $c_1, ..., c_\alpha$ and $x_i = 1$, $i \in \overline{1, \alpha}$, $x_k = 0$, $k \in \overline{\alpha + 1, n}$. $p^* = \sum\limits_{i=1}^{\alpha} c_i$.

$\nu^* = -c_\alpha$; $\lambda^*_{1_i} = 0, i \in \overline{1, \alpha}$; $\lambda_{1_i} = -(c_i + c_\alpha), i \in \overline{\alpha + 1, n}$; $\lambda^*_{2_i} = 0, i \in \overline{1, \alpha}$; $\lambda^*_{2_i} = c_i - c_\alpha, i \in \overline{\alpha + 1, n}$; for such $\nu^*, \lambda^*_1, \lambda^*_2$ all kkt is right.

If α is not integer, then we take like first $\lfloor \alpha \rfloor$ elements and difference between α and $\lfloor \alpha \rfloor$.

$x_i = 1$, $i \in \overline{1, \lfloor \alpha \rfloor - 1}$, $x_{\lfloor \alpha \rfloor} = 1 + \alpha - \lfloor \alpha \rfloor$, $x_k = 0$, $k \in \overline{\lfloor \alpha \rfloor + 1, n}$. $p^* = \sum\limits_{i=1}^{\lfloor \alpha \rfloor - 1} c_i + (1 + \alpha - \lfloor \alpha \rfloor) \cdot c_{\lfloor \alpha \rfloor}$. $\nu^*, \lambda^*_1, \lambda^*_2$ we take from previos task.

If $\mathbb{1}^T x \le \alpha$, then we can take first $k \le \alpha$ elements in c such that $c_i < 0$, $i \in \overline{1, k}$, $0 \le c_{k+1}$. And $p^* = \sum\limits_{i=1}^{k} c_i$. If all components are positive then we tak $x^* = 0$. $k \ge \alpha$ we take first alpha components if they are negative. And $x^*_i = 1, i \in \overline{1, k}$; $x^*_i = 0, i \in \overline{k + 1, n}$

Answer: $x^*, \nu^*, \lambda^*_1, \lambda^*_2$ are described above.

$$\begin{cases} p^* = \sum\limits_{i=1}^{\alpha} c_i, \alpha \text{ is an integer} \\ p^* = \sum\limits_{i=1}^{\lfloor \alpha \rfloor} c_i + (\alpha - \lfloor \alpha \rfloor) \cdot c_{\lfloor \alpha \rfloor + 1}, \alpha \text{ is not an integer} \\ p^* = \sum\limits_{i=1}^{k} c_i \text{ } , \mathbb{1}^T x \le \alpha, \text{the conditions for k are given above (if negatives components exist} \end{cases}$$

8.5 Problem №5

Give an explicit solution of the following QP.

$$c^T x \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } x^T Ax \leq 1$$

where $A \in \mathbb{S}^n_{++}, c \neq 0$. What is the solution if the problem is not convex ($A \notin \mathbb{S}^n_{++}$). (Hint: consider eigendecomposition of the matrix: $A = Qdiag(\lambda)Q^T = \sum\limits_{i=1}^{n} \lambda_i q_i q_i^T$ and different cases of $\lambda > 0, \lambda = 0, \lambda < 0$)?

Solution: Lagrangian:

$$L(x, \lambda) = c^T x + \lambda(x^T Ax - 1)$$

KKT conditions:

$$\begin{cases} \nabla_x L = c + (A + A^T)x = c + 2\lambda Ax = 0 \\ \lambda \geq 0 \\ \lambda(x^T Ax - 1) = 0 \\ x^T Ax \leq 1 \end{cases}$$

KKT conditions will be necessary and sufficient, because our problem is convex and Slater's conditions are met. It's right, we can take x such as first component will be equals one divide on ten multiply sum of components of first line of matrix A, and others components of x will be equal zero. From that we get that for such x: $x^T Ax = \frac{1}{10} < 1$. And Slater's conditions are met.

$$x = -\frac{A^{-1}c}{2\lambda}$$

$$\lambda(\frac{c^T A^{-1}c}{4\lambda^2} - 1) = 0$$

$$\lambda = \sqrt{\frac{c^T A^{-1}c}{4}}$$

And x^* will be:

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

And optimal value will be: $p^* = -\sqrt{c^T A{-}1c}$

If matrix is not positive symmertrical.

If $\lambda_k > 0$, $i \in \overline{1, n}$, then it's the same task.

If $\lambda_k \leq 0$, then we can take $y_k \to \pm\infty$, and minimum will be $-\infty$.

If $\lambda_k = 0$, if $b_k = 0$ for all k such $\lambda_k = 0$, then it's the same task, but with less $\lambda_l > 0$.

Answer: $p^* = -\sqrt{c^T A^{-1}c}$, $x^* =$

8.6 Problem №6

Give an explicit solution of the following QP.

$$c^T x \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } (x - x_c)^T A (x - x_c) \leq 1$$

where $A \in \mathbb{S}^n_{++}, c \neq 0, x_c$

Solution: We will take $y = (x - x_c)$, and write answer, because we can rewrite it as the fifth problem we solved.

$$y^* = -\frac{A^{-1}c}{\sqrt{c^T A^{-1} c}}, \, x^* = x_c - \frac{A^{-1}c}{\sqrt{c^T A^{-1} c}}$$

$$p^* = -\sqrt{c^T A^{-1} c} + c^T x_c$$

Answer: $p^* = -\sqrt{c^T A^{-1} c} + c^T x_c$, $x^* = x_c - \frac{A^{-1}c}{\sqrt{c^T A^{-1} c}}$

8.7 Problem №7

Give an explicit solution of the following QP.

$$x^T B x \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } x^T A x \leq 1$$

where $A \in \mathbb{S}^n_{++}, B \in \mathbb{S}^n_{+}$

Solution: We know that $B \in \mathbb{S}^n_+$, and from that $0 \leq x^T B x$ for any $x \in \mathbb{R}^n$, and minimum will be when $x^* = 0$, let's show that x^* is feasible for our conditions $(x^*)^T A x^* = 0 \leq 1$. Wohoo!

Answer: $p^* = 0$, $x^* = 0$.

8.8 Problem №8

Consider the equality constrained least-squares problem

$$||Ax - b||_2^2 \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } Cx = d$$

where $A \in \mathbb{R}^{m\dot{n}}$ with

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

Solution: If system $Cx = d$ has no solutions, then $p^* = +\infty$

If system $Cx = d$ has solutions, then by Slater's conditions we have an internal point:

$$L(x, \nu) = \langle Ax - b, Ax - b\rangle + \nu^T(Cx - d)$$

$$\begin{cases} \nabla_x L = 2A^T(Ax - b) + \nu^T C = 0 \\ \nabla_\nu L = Cx - d = 0 \end{cases}$$

$$\begin{cases} x = (A^T A)^\dagger(A^T b - \frac{1}{2}\nu^T C) \\ x = C^\dagger d \end{cases}$$

Then we get:

$$(A^T A)^\dagger(A^T b - \frac{1}{2}\nu^T C) = C^\dagger d$$

$$A^T b - \frac{1}{2}\nu^T C = (A^T A)^\dagger)^\dagger(C^\dagger d) = (A^T A)(C^\dagger d)$$

Then we get ν^*

$$\nu^* = 2(A^T bC^\dagger - (A^T A)(C^\dagger d)C^\dagger)^T$$

Answer:

$$\begin{cases} p^* = +\infty, \text{ if system } Cx = d \text{ has no solutions} \\ x^* = C^\dagger d, \nu^* = 2(A^T bC^\dagger - (A^T A)(C^\dagger b)C^\dagger)^T, \text{ otherwise} \end{cases}$$

8.9 Problem №9

Derive the KKT conditions for the problem

$$trX - \log(detX) \to \min_{X \in \mathbb{S}^n_{++}}$$

$$\text{s.t. } Xs = y$$

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

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

Solution: Let's write Lagrangian:

$$L(x, \nu) = trX - \log(detX) + \nu^T(Xs - y)$$

It can be seen that Slaters conditions are met: problem is convex and all functions in conditions are affine.

$$\begin{cases} \nabla_x L = I - X^{-1} + \nu s^T = 0 \mid \text{on right } \cdot y \\ \nabla_\nu L = Xs - y = 0 \end{cases}$$

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

$$X^* s - y = Is + yy^T s - \frac{1}{s^T s} ss^T s - y = s + y - s - y = 0$$

Wohoo, it's feasible for KKT! Let's find nu for that point:

$$\begin{cases} y - X^{-1} y + \nu s^T y = 0 \\ s = X^{-1} y \end{cases}$$

$$\begin{cases} y - s + \nu s^T y = 0 \mid \text{on left } \cdot y^T \\ y^T y - y^T s + y^T \nu s^T y = y^T y - 1 + y^T \nu = 0 \mid \text{transpose} \end{cases}$$

$$\begin{cases} \nu y^T = 1 - y^T y \mid \text{on right } \cdot s \\ \nu = s - y^T y s \end{cases}$$

We can find X^{-1} from first condition of KKT:

$$X^{-1} = I + (s - y^T y s) s^T = I + ss^T - y^T y s s^T$$

Let's check $X^* X^{-1} = I$

$$(I + yy^T - \frac{ss^T}{s^T s})(I + ss^T - y^T y s s^T) = I + ss^T - y^T y s s^T + +yy^T + yy^T ss^T - yy^T y^T y s s^T$$
$$- \frac{ss^T}{s^T s} - \frac{ss^T ss^T}{s^T s} + \frac{ss^T}{ss^T} y^T y s s^T = I$$

Check, that X lie in $\mathbb{S}^n_{++}$ is very easy: $I + yy^T - \frac{1}{s^T s} ss^T$ is positive and symmetric matrix. I want to say that is easy you need to look at matrix and understand, that $I - \frac{1}{s^T s} ss^T$ is symmetric and positive and yy^T is also positive and symmetric, but some will people can not understand that. Okay, let's do it by definition:

$$x^T (I + yy^T - \frac{1}{s^T s} ss^T) x = (x^T + x^T yy^T - \frac{x^T}{s^T s} ss^T) x = ...$$

$$... = x^T x + (y^T x)^T (y^T x) - \frac{1}{s^T s} (s^T x)^T (s^T x) = \langle x, x \rangle + \langle y, x \rangle^2 - \frac{\langle s, x \rangle^2}{s^T s} \geq 0$$

It's right because length of projection on unit vector is less then length of vector in space.

8.10 Problem №10

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

$$f_0(x) \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } f_i(x) \leq 0, \; i = \overline{1, m}$$

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

$$\begin{cases} \nabla_x L(x^*, \mu^*) = \nabla f_0(x^*) + \sum\limits_{i=1}^{m} \mu_i^* \nabla f_i(x^*) = 0 \\ \mu_i^* \geq 0, \; i = \overline{1, m} \\ \mu_i^* f_i(x^*) = 0, \; i = \overline{1, n} \\ f_i(x^*) \leq 0, \; i = \overline{1, n} \end{cases}$$

Show that:

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

for all feasible x. In other words the KKT conditions imply the simple optimality criterion or $\nabla f_0(x^*)$ defines a supporting hypperplane to the feasible set at x^*.

Solution: We all know that for convex functions it is right that a criterion for feasible x^*:

$$f_i(x) - f_i(x^*) \geq \nabla f_i(x^*)^T (x - x^*)$$

Let's multiply it by μ_i^* and sum over i:

$$\sum_{i=1}^{m} f_i(x) \mu_i^* \geq \sum_{i=1}^{m} f_i(x^*) \mu_i^* + \sum_{i=1}^{n} \mu_i^* \nabla f_i(x^*)^T (x - x^*)$$

From KKT we know: $\sum\limits_{i=1}^{n} \mu_i^* \nabla f_i(x^*)^T (x - x^*) = -\nabla f_0(x^*)^T (x - x^*)$, $f_i(x) \leq 0$ and $f_i^*(x^*) \mu_i^* = 0$.

$$0 \leq \sum_{i=1}^{m} f_i(x) \mu_i^* \leq -\nabla f_0(x^*)^T (x - x^*)$$

And we get:

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

9 Duality

9.1 Problem №1

Express the dual problem of

$$c^T x \to 1 \min_{x \in \mathbb{R}^n}$$
$$\text{s.t. } f(x) \leq 0$$

with $c \neq 0$, in terms of the conjugate function f^*. Explain why the problem you give is convex. We do not assume f is convex.

Solution: Let's write lagrangian ($\lambda \geq 0, \lambda \in \mathbb{R}^n$):

$$L(x, \lambda) = c^T x + \lambda f(x)$$

$$g(\lambda) = \inf_{x \in \mathbb{R}^n} (c^T x + \lambda f(x)) = -\lambda \sup_{x \in \mathbb{R}^n} (-\frac{1}{\lambda} c^T x - f(x)) = \mid \text{by definition} \mid = -\lambda f^*(-\frac{c}{\lambda})$$

Okay, let's write dual problem.

$$-\lambda f^*(-\frac{c}{\lambda}) \to \max_{\lambda \in \mathbb{R}}$$
$$\text{s. t. } \lambda \geq 0$$

Answer: It's concave function by definition of conjugate function.

9.2 Problem №2

Minimum volume covering ellipsoid. Let we have the primal problem:

$$\log \det X^{-1} \to \min_{X \in \mathbb{S}^n_{++}}$$

$$\text{s. t. } a_i^T X a_i \leq 1, \ i = 1, ..., m$$

1. Find Lagrangian of the primal problem
2. Find the dual function
3. Write down the dual problem
4. Check whether problem holds strong duality or not
5. Write down the solution of the dual problem

Solution: $f_i(X) = a_i^T X a_i - 1, \ f(X) = (f_1, ..., f_m)$.

1.
$$L(x, \lambda) = \log \det X^{-1} + \sum_{i=1}^{m} \lambda_i (a_i^T X a_i - 1) = \log \det X^{-1} + \lambda^T f(X)$$

2. We can really easilly find the dual function for this problem because we have already found the conjugate function in the first homework.

$$f^*(Y) = -n + \log\det(-Y^{-1}), \text{ where } Y \in \mathbb{S}^n_{++}$$

$$g(\lambda, \nu) = -b^T\lambda - d^T\nu - f_0^*(-A^T\lambda - C^T\nu)$$

$$g(\lambda) = \begin{cases} \log\det\left(\sum\limits_{i=1}^m \lambda_i a_i a_i^T\right) - \mathbb{1}^T\lambda + n, \sum\limits_{i=1}^m \lambda_i a_i a_i^T \succ 0 \\ +\infty, \text{ otherwise} \end{cases}$$

3. $$\log\det\left(\sum_{i=1}^m (\lambda_i a_i a_i^T)\right) + n - \lambda^T\mathbb{1} \to \max_{\lambda\in\mathbb{R}^m}$$
$$\text{s. t. } \lambda \succcurlyeq 0$$

4. Strong duality holds because Slater's conditions are performed, it means that $\exists X \in \mathbb{S}^n_{++}$ such that $a_i^T X a_i < 1, i \in \overline{1,m}$. For example $X = diag(\frac{1}{10a_i^Ta_i}, ..., \frac{1}{2a_i^Ta_i})$

5. Let's solve it! From fully symmetric problem and many iterations, when I am trying to solve this problem, I get that λ will be like that $\lambda = (\alpha, \alpha, ..., \alpha)$.

   Okay, it's feasible to our budget set, let's put in our function.

$$g(\lambda) = \log\det\left(\sum_{i=1}^m (\alpha a_i a_i^T)\right) + n - \alpha m = \log\alpha^n \det\left(\sum_{i=1}^m a_i a_i^T\right) + n - \alpha m$$

   Okay, from it we need to maximize only: $h(\alpha) = n\log\alpha - \alpha m$, it's concave function and local maximum will be global.

$$h'(\alpha) = \frac{n}{\alpha} - m = 0$$

   From it we get that: $\alpha^* = \frac{n}{m}$ and $h(\alpha^*) = n\log\frac{n}{m} - n$

$$g(\lambda^*) = \log\det\left(\sum_{i=1}^m a_i a_i^T\right) + n\log\frac{n}{m}$$

   We all remember, that:

$$\nabla_X L(X, \lambda) = -X^{-1} + \sum_{i=1}^m \lambda_i a_i a_i^T = 0$$

$$\nabla_\lambda L(X, \lambda) = \sum_{i=1} (a_i^T X a_i - 1) = 0$$

Okay, let's show that $\nabla_\lambda L(X^*, \lambda^*) = 0$, if $X^-1 = \sum\limits_{i=1}^{m} \lambda_i a_i a_i^T$.

$$\nabla_\lambda L \sum_{i=1} (a_i^T \left(\sum_{i=1}^{m} \alpha a_i a_i^T \right)^{-1} a_i) - m = \sum_{i=1}^{m} \langle a_i a_i^T, \left(\sum_{i=1}^{m} \alpha a_i a_i^T \right)^{-1} \rangle - m$$

And we get:

$$\langle \sum_{i=1}^{m} a_i a_i^T, \left(\sum_{i=1}^{m} \alpha a_i a_i^T \right)^{-1} \rangle - m = \frac{n}{\alpha} - m = \ | \ \alpha = \frac{n}{m} | \ = \frac{n}{n/m} - m = 0$$

Youhoo, our α^* is feasible for Slater's conditions. And if we input X^* into $f_0(X)$ we will get:

$$f_0(X^*) = \log\det(\sum_{i=1}^{m} \alpha a_i a_i^T) = n \log \alpha + \log\det \sum_{i=1}^{m} a_i a_i^T = n \log \frac{n}{m} + \log\det \sum_{i=1}^{m} a_i a_i^T$$

$$g(\lambda^*) = \log\det \sum_{i=1}^{m} a_i a_i^T + n \log \frac{n}{m}$$

We show strong duality and solve duality problem.

9.3 Problem №3

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

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

where $f_0(x) : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable, and $A \in \mathbb{R}^{m\dot{n}}$ with $\text{rank} A = m$.

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, $\widetilde{x}$, should be an approximate solution of the original problem. Intuition suggests that the larger the penalty weight α, the better the approximation $\widetilde{x}$ to a solution of the original problem. Suppose $\widetilde{x}$ is a minimizer of $\phi(x)$. Show how to find, from $\widetilde{x}$, a dual feasible point for the original problem. Find the corresponding lower bound on the optimal value of the original problem.

Solution: Let's write lagrangian:

$$L(x, \lambda) = f_0(x) + \lambda^T(Ax - b)$$

$$\nabla_x L(x, \lambda) = \nabla f_0(x) + A^T \lambda$$

If $\widetilde{x}$ is min in $\phi(x)$, then we get:

$$\nabla \phi(\widetilde{x}) = \nabla f_0(\widetilde{x}) + 2\alpha A^T(A\widetilde{x} - b) = 0$$

And from that we get:

$$\lambda = 2\alpha(A\widetilde{x} - b)$$

Let's write the dual problem to that:

$$g(\lambda) = \inf_{x \in \mathbb{R}^n} (f_0(x) + \lambda^T(Ax - b) = f_0(\widetilde{x}) + 2\alpha||A\widetilde{x} - b|_2^2$$

And for all x such that: $Ax = b$

$$f_0(x) \geq f_0(\widetilde{x}) + 2\alpha||A\widetilde{x} - b||_2^2$$

And when we increasing alpha we get higher lower bound for the original problem.

9.4 Problem №4

Analytic centering. Derive a dual problem for

$$-\sum_{i=1}^{m} \log(b_i - a_i^T x) \to \min_{x \in \mathbb{R}^n}$$

with domain $\{x | a_i^T x < b_i, i \in \overline{1, m}\}$

First introduce new variables Z_i and equality constraints $Z_i = b_i - a_i^T x$. (The solution of this problem is called the analytic center of the linear inequalities $a_i^T x \leq b_i$, $i \in \overline{1, m}$. Analytic centers have geometric applications, and play an important role in berrier methods).

Solution: Let's take $Z_i = b_i - a_i^T x$.

$$L(x, Z, \lambda) = -\sum_{i=1}^{m} \log Z_i + \sum_{i=1}^{m} \lambda_i(Z_i - b_i + a_i^T x)$$

The dual function will be:

$$g(\lambda) = \inf_{x,Z \in domain} (-\sum_{i=1}^{m} \log Z_i + \sum_{i=1}^{m} \lambda_i (Z_i + a_i^T x)) + \lambda^T b$$

If $\sum\limits_{i=1}^{m} \lambda_i a_i^T = \lambda^T A = 0$ it's not true, than it will unbounded bellow. If $\lambda \not\succ 0$ it will be also unbounded below. Let's write our function, the optimal is $Z_i = \frac{1}{\lambda_i}$.

$$g(\nu) = \begin{cases} \sum\limits_{i=1}^{m} \log \lambda_i + m - \lambda^T b, \ \lambda^T A = 0, \ \lambda \succ 0 \\ -\infty, \text{ otherwise} \end{cases}$$

Dual problem will be:

$$\sum_{i=1}^{m} \log \lambda_i + m - \lambda^T b \to \max_{\lambda \in \mathbb{R}^m}$$

$$\text{s.t. } \lambda^T A = 0$$
$$\lambda \succ 0$$

10 Linear Programming

10.1 Problem №1

The company's production facilities are such that if we devote the entire production to headphones covers, we can produce 5000 of them in one day. If we devote the entire production to phone covers or laptop covers, we can produce 4000 or 2000 of them in one day.

The production schedule is one week (6 working days), and the week's production must be stored before distribution. Storing 1000 headphones covers (packaging included) takes up 30 cubic feet of space. Storing 1000 phone covers (packaging included) takes up 50 cubic feet of space, and storing 1000 laptop covers (packaging included) takes up 220 cubic feet of space. The total storage space available is 1500 cubic feet.

Due to commercial agreements with Random Corp has to deliver at least 4500 headphones covers and 3000 laptop covers per week in order to strengthen the product's diffusion.

The marketing department estimates that the weekly demand for headphones covers, phone, and laptop covers does not exceed 9000 and 14000, and 7000 units, therefore the company does not want to produce more than these amounts for headphones, phone, and laptop covers.

Finally, the net profit per each headphones cover, phone cover, and laptop cover is 5, 7 and 12 \$, respectively.

The aim is to determine a weekly production schedule that maximizes the total net profit.

a. Write a Linear Programming formulation for the problem. Use following variables: y_1, y_2, y_3 - number of covers for headpnones, phones and laptops produced per week.

b. Finf the solution to the problem using PyOMO.

Solution:

$$\begin{bmatrix}5, 7, 12\end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \to \max_{y \in \mathbb{R}^3}$$

$$\begin{aligned}
s.t. \left(y_1 \frac{30}{1000} + y_2 \frac{50}{1000} + y_3 \frac{220}{1000} \right) &\leq 1500 \\
4500 \leq y_1 &\leq 9000 \\
0 \leq y_2 &\leq 14000 \\
3000 \leq y_3 &\leq 7000 \\
\frac{y_1}{5000} + \frac{y_2}{4000} + \frac{y_3}{2000} &\leq 6
\end{aligned}$$

Let's rewrite this problem in terms of x: $y_1 = x_1 + 4500$, $y_2 = x_2$, $y_3 = x_3 + 3000$.

$$[5, 7, 12] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + 58500 \to \max_{x \in \mathbb{R}^3}$$

$$s.t. \left(x_1 \frac{30}{1000} + x_2 \frac{50}{1000} + x_3 \frac{220}{1000} \right) \leq 705$$

$$0 \leq x_1 \leq 4500$$

$$0 \leq x_2 \leq 14000$$

$$0 \leq x_3 \leq 4000$$

$$\frac{x_1}{5000} + \frac{x_2}{4000} + \frac{x_3}{2000} \leq 3.6$$

LP_covery.dat

```
param:  F:                 c     V  :=
  "Headphones"             5.0    0.0002
  "Phones"                 7.0    0.00025
  "Laptops"                12.0   0.0005;

param Vmax := 3.5;

param:  N:       Nmin   Nmax :=
        Size      0      1500
        lim1      4500   9000
        lim2      0      14000
        lim3      3000   7000;

param a:
                             Size   lim1    lim2   lim3 :=
  "Headphones"               0.03    1.0     0.0    0.0
  "Phones"                   0.05    0.0     1.0    0.0
  "Laptops"                  0.22    0.0     0.0    1.0    ;
```

data for linear programming in terms of y

LP_cover.dat

```
param:  F:                 c     V  :=
  "Headphones"             5.0    0.0002
  "Phones"                 7.0    0.00025
  "Laptops"                12.0   0.0005;

param Vmax := 3.5;

param:  N:       Nmin   Nmax :=
        Size      0      705
        lim1      0      4500
        lim2      0      14000
        lim3      0      4000;

param a:
                             Size   lim1    lim2   lim3 :=
  "Headphones"               0.03    1.0     0.0    0.0
  "Phones"                   0.05    0.0     1.0    0.0
  "Laptops"                  0.22    0.0     0.0    1.0    ;

```

data for linear programming in terms of x

```python
# LP_covery.py
from pyomo.environ import *
infinity = float('inf')

model = AbstractModel(sense=pyo.maximize)
# Covers
model.F = Set()

# Storage
model.N = Set()

# Revenue of each cover
model.c = Param(model.F, within=PositiveReals)
# Size of each cover
model.a = Param(model.F, model.N, within=NonNegativeReals)

# Lower and upper bound on each cover
model.Nmin = Param(model.N, within=NonNegativeReals, default =0.0)
model.Nmax = Param(model.N, within=NonNegativeReals, default=infinity)

# Maximum producing per each cover
model.V = Param(model.F, within=PositiveReals)

# Maximum producing of cover consumed
model.Vmax = Param(within=PositiveReals)

# Number of cover
model.y = Var(model.F, within=NonNegativeIntegers)

# Maximize the revenue of covered
def cost_rule(model):
    return sum(model.c[i] * model.y[i] for i in model.F)

# Limit for volume of covers
def volume_rule(model, j):
    value = sum(model.a[i, j] * model.y[i] for i in model.F)
    return inequality(model.Nmin[j], value, model.Nmax[j])

model.volume_limit = Constraint(model.N, rule=volume_rule)

def producing_rule(model):
    return sum(model.V[i]*model.x[i] for i in model.F) <= model.Vmax

model.producing = Constraint(rule=producing_rule)
```

solutions in terms of y

```
# ----------------------------------------------------------
#   Solution Information
# ----------------------------------------------------------
Solution:
- number of solutions: 1
  number of solutions displayed: 1
- Gap: 0.0
  Status: optimal
  Message: None
  Objective:
    cost:
      Value: 158400
  Variable:
    y[Headphones]:
      Value: 6000
    y[Laptops]:
      Value: 3000
    y[Phones]:
      Value: 13200
  Constraint: No values
```

answer problem in terms of y

LP_covery.py

```python
from pyomo.environ import *
infinity = float('inf')

model = AbstractModel(sense=pyo.maximize)
# Covers
model.F = Set()

# Storage
model.N = Set()

# Revenue of each cover
model.c = Param(model.F, within=PositiveReals)
# Size of each cover
model.a = Param(model.F, model.N, within=NonNegativeReals)

# Lower and upper bound on each cover
model.Nmin = Param(model.N, within=NonNegativeReals, default =0.0)
model.Nmax = Param(model.N, within=NonNegativeReals, default=infinity)

# Maximum producing per each cover
model.V = Param(model.F, within=PositiveReals)

# Maximum producing of cover consumed
model.Vmax = Param(within=PositiveReals)

# Number of cover
model.y = Var(model.F, within=NonNegativeIntegers)

# Maximize the revenue of covered
def cost_rule(model):
    return sum(model.c[i] * model.y[i] for i in model.F)

# Limit for volume of covers
def volume_rule(model, j):
    value = sum(model.a[i, j] * model.y[i] for i in model.F)
    return inequality(model.Nmin[j], value, model.Nmax[j])

model.volume_limit = Constraint(model.N, rule=volume_rule)

def producing_rule(model):
    return sum(model.V[i]*model.x[i] for i in model.F) <= model.Vmax

model.producing = Constraint(rule=producing_rule)
```

solutions in terms of x

```
# ----------------------------------------------------------
#   Solution Information
# ----------------------------------------------------------
Solution:
- number of solutions: 1
  number of solutions displayed: 1
- Gap: 0.0
  Status: optimal
  Message: None
  Objective:
    cost:
      Value: 158400
  Variable:
    x[Headphones]:
      Value: 1500
    x[Phones]:
      Value: 13200
  Constraint: No values
```

answer problem in terms of x

10.2 Problem №2

Prove the optimality of the solution

$$x^T = \left(\frac{5}{26}, \frac{5}{2}, \frac{27}{26}\right)$$

to the following linear programming problem:

$$9x_1 + 14x_2 + 7x_3 \to \max_{x \in \mathbb{R}^n}$$

$$\begin{aligned} \text{s.t. } 2x_1 + x_2 + 3x_3 &\le 6 \\ 5x_1 + 4x_2 + x_3 &\le 12 \\ 2x_2 &\le 5 \end{aligned}$$

buy you cannot use any numerical algorithm here.

Solution: Point $x = (0,0,0)^T$ satisfies Slater's conditions.

$$\begin{gathered} 0 < 6 \\ 0 < 12 \\ 0 < 5 \end{gathered}$$

And KKT becomes necessary and sufficient.

$$L(x, \lambda) = 9x_1 + 14x_2 + 7x_3 - \lambda_1(2x_1 + x_2 + 3x_3 - 6) - \lambda_2(5x_1 + 4x_2 + x_3 - 12) - \lambda_3(2x_2 - 5)$$

$$\begin{cases} \nabla_{x_1} L = 9 - 2\lambda_1 - 5\lambda_2 = 0 \\ \nabla_{x_2} L = 14 - \lambda_1 - 4\lambda_2 - 2\lambda_3 = 0 \\ \nabla_{x_3} L = 7 - 3\lambda_1 - \lambda_2 = 0 \\ \lambda \succcurlyeq 0 \\ \lambda_1(2x_1 + x_2 + 3x_3 - 6) = 0 \\ \lambda_2(5x_1 + 4x_2 + x_3 - 12) = 0 \\ \lambda_3(2x_2 - 5) = 0 \\ 2x_1 + x_2 + 3x_3 \le 6 \\ 5x_1 + 4x_2 + x_3 \le 12 \\ 2x_2 \le 5 \end{cases}$$

From it we get:

$$\begin{cases} \lambda_1 = 2 \\ \lambda_2 = 1 \\ \lambda_3 = 4 \\ \lambda \succcurlyeq 0 \\ 2 \cdot (2 \cdot \frac{5}{26} + \frac{5}{2} + 3 \cdot \frac{27}{27} - 6) = 2 \cdot (6 - 6) = 0 \\ 1 \cdot (5 \cdot \frac{5}{26} + 4 \cdot \frac{5}{2} + \frac{27}{27} - 12) = 12 - 12 = 0 \\ 4 \cdot (2\frac{5}{2} - 5) = 0 \\ 2 \cdot \frac{5}{26} + \frac{5}{2} + 3 \cdot \frac{27}{27} = 6 \leq 6 \\ 5 \cdot \frac{5}{26} + 4 \cdot \frac{5}{2} + \frac{27}{27} = 12 \leq 12 \\ 2\frac{5}{2} = 5 \leq 5 \end{cases}$$

$L(\dfrac{5}{26}, \dfrac{5}{2}, \dfrac{27}{26}) = 9 \cdot \dfrac{5}{26} + 14\dfrac{5}{2} + 7\dfrac{27}{26} - 2 \cdot (6-6) - (12-12) - 4 \cdot (5-5) = \dfrac{45 + 70 + 189}{26} = \dfrac{304}{26}$

Optimal point: $\frac{152}{13}$

We show that this problem satisfies Slater's conditions and KKT becomes essential and suffiecient condition. And point $x = \left(\frac{5}{26}, \frac{5}{2}, \frac{27}{26}\right)^T$ satisfies the conditions of the KKT.

10.3 Problem №3

Transform the following linear program into an equivalent linear programm in standart form($c^T x \to \max_{x \in \mathbb{R}^n} : Ax = b, x \geq 0$):

$$x_1 - x_2 \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } 2x_1 + x_2 \geq 3$$
$$3x_1 - x_2 \leq 7$$
$$x_1 \geq 0$$

Solution:

Let's take $x_1 = x$, $X_2 = y - z$, where $y = x_{2+}$, $z = x_{2-}$, $y, z \geq 0$ And rewrite our problem.

$$x_1 - x_2 \to \min_{x \in \mathbb{R}^3}$$

$$\text{s.t.} \quad -2x_1 - y + z \leq 3$$
$$3x_1 - y + z \leq 7$$
$$x_1 \geq 0$$
$$y \geq 0$$
$$z \geq 0$$

Okay, let's again rewrite it.

$$x - y + z \to \min_{x \in \mathbb{R}^3}$$
$$\text{s.t.} \quad \begin{pmatrix} 2 & 1 & -1 \\ 3 & -1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \succcurlyeq \begin{pmatrix} 3 \\ 3 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

But we need to rewrite it standart form, not it's canonical form.

$$c = (1, -1, 0)^T$$
$$A = \begin{pmatrix} 2 & 1 & -1 \\ 3 & -1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$
$$b = (3, 3, 0, 0, 0)^T$$

$$L(x, \lambda) = c^T x + \lambda^T (Ax - b) = (c^T + \lambda^T A -)x - \lambda^T b$$
$$g(\lambda) = \inf_{x \in \mathbb{R}^3} \left(c^T + \lambda^T A\right) x - \lambda^T b$$

And we have:

$$-b^T \lambda \to \max_{\lambda \in \mathbb{R}^5}$$
$$\text{s.t.} \ A^T \lambda = -c$$
$$\lambda \succcurlyeq 0$$

And finally:

$$-\left(3, 3, 0, 0, 0\right)^T \lambda \to \max_{\lambda \in \mathbb{R}^5}$$
$$\text{s.t.} \quad \begin{pmatrix} 2 & 3 & -1 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 & -1 \end{pmatrix} \lambda = (-1, 1, 0)^T$$
$$\lambda \succcurlyeq 0$$

10.4 Problem №4

Consider

$$4x_1 + 5x_2 + 2x_3 \to \max_{x \in \mathbb{R}^3}$$
$$\text{s.t. } 2x_1 - x_2 + 2x_3 \leq 9$$
$$3x_1 + 5x_2 + 4x_3 \leq 8$$
$$x_1 + x_2 + 2x_3 \leq 2$$
$$x_1, x_2, x_3 \geq 0$$

a Find an optimal solution to the Linear Programming using the simplex method.

b Write the dual linear program. Find an optimal dual solution. Do we have strong duality here?

Solution: Let's rewrite problem to normal canonical form (by scheme from lecture):

$$4x_1 + 5x_2 + 2x_3 - x_7 - x_8 - x_9 \to \max_{x \in \mathbb{R}^9}$$
$$\text{s.t. } 2x_1 - x_2 + 2x_3 + x_4 + x_7 = 9$$
$$3x_1 + 5x_2 + 4x_3 + 4x_4 + x_5 + x_8 = 8$$
$$x_1 + x_2 + x_3 + x_6 + x_9 = 2$$
$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9 \geq 0$$

$4x_1 + 5x_2 + 5x$

First table of simplex method:

	c*		0	0	0	0	0	0	-1	-1	-1	
Basis		b	a1	a2	a3	a4	a5	a6	a7	a8	a9	t
a7	-1	9	2	-1	2	1	0	0	1	0	0	4.5
a8	-1	8	3	5	4	0	1	0	0	1	0	2
a9	-1	2	1	1	2	0	0	1	0	0	1	1
z		-19	-6	-5	-8	-1	-1	-1	-1	-1	-1	
delta			-6	-5	-8	-1	-1	-1	0	0	0	

New basis: a_3, a_7, a_8.

	c*		0	0	0	0	0	0	-1	-1	-1	
Basis		b	a1	a2	a3	a4	a5	a6	a7	a8	a9	t
a3	0	1	0.5	0.5	1	0	0	0.5	0	0	0.5	2
a7	-1	7	1	-2	0	1	0	-1	1	0	-1	7
a8	-1	4	1	3	0	0	1	2	0	1	2	4
z		-11	-2	-1	0	-1	-1	-1	-1	-1	-1	
delta			-2	-1	0	-1	-1	-1	0	0	0	

New basis: a_1, a_7, a_8

	c*		0	0	0	0	0	0	-1	-1	-1	
Basis		b	a1	a2	a3	a4	a5	a6	a7	a8	a9	t
a1	0	2	1	1	2	0	0	1	0	0	1	—
a7	-1	5	0	-3	-2	1	0	-2	1	0	-2	5
a8	-1	2	0	2	2	0	1	-3	0	1	-3	—
z		-7	0	1	0	-1	-1	5	-1	-1	5	
delta			0	1	0	-1	-1	5	0	0	6	

New basis: a_1, a_7, a_5.

	c*		0	0	0	0	0	0	-1	-1	-1	
Basis		b	a1	a2	a3	a4	a5	a6	a7	a8	a9	t
a1	0	2	1	1	2	0	0	1	0	0	1	–
a7	-1	5	0	-3	-2	1	0	-2	1	0	-2	5
a5	0	2	0	2	2	0	1	-3	0	1	-3	–
z		-5	0	3	2	-1	0	2	-1	0	2	
delta			0	3	2	-1	0	2	0	1	3	

New basis: a_1, a_4, a_5.

	c*		0	0	0	0	0	0	-1	-1	-1	
Basis		b	a1	a2	a3	a4	a5	a6	a7	a8	a9	t
a1	0	2	1	1	2	0	0	1	0	0	1	
a4	0	5	0	-3	-2	1	0	-2	1	0	-2	
a5	0	2	0	2	-2	0	1	-3	0	1	-3	
z		0	0	0	0	0	0	0	0	0	0	
delta			0	0	0	0	0	0	1	1	1	

$\Delta \geq 0$, from that we get $x = (2, 0, 0, 5, 2, 0, 0, 0, 0)^T$ - solution of this problem.
Extreme point for the original problem will be: $x = (2, 0, 0, 5, 2, 0)^T$.
Solution of the original problem.

	c*		4	5	2	0	0	0	
Basis		b	a1	a2	a3	a4	a5	a6	t
a1	4	2	1	1	2	0	0	1	2
a4	0	5	0	-3	-2	1	0	-2	
a5	0	2	0	2	2	0	1	-3	1
z		8	4	4	8	0	0	4	
delta			0	-1	6	0	0	4	

New basis: a_1, a_2, a_5.

	c*		4	5	2	0	0	0	
Basis		b	a1	a2	a3	a4	a5	a6	t
a1	4	1	1	0	3	0	-0.5	2.5	
a2	5	1	0	1	-1	0	0.5	-1.5	
a4	0	8	0	0	-5	1	1.5	-6.5	
z		9	4	5	7	0	0.5	2.5	
delta			0	0	5	0	0.5	2.5	

Solution is $(1, 1, 0, 8, 0, 0)^T$. Comeback to our original problem, get rid of 3 latest coordinates, point will be: $(1, 1, 0)^T$ on that point maximum will be achieve. Maximum equals 9.

Answer: maximum equals 9, in point $x = (1, 1, 0)^T$.

Solution:

$$L(x, \lambda) = 4x_1 + 5x_2 + 2x_3 - \lambda_1(2x_1 - x_2 + 2x_3 - 9) - \lambda_2(3x_1 + 5x_2 + 4x_3 - 8) \\ -\lambda_3(x_1 + x_2 + 2x_3 - 2) + \lambda_4 x_1 + \lambda_5 x_2 + \lambda_6 x_3$$

Slater's conditins met and KKT conditions are necessary and sufficient.

$$\begin{cases} \nabla_{x_1} L = 4 - 2\lambda_1 - 3\lambda_2 - \lambda_3 + \lambda_4 = 0 \\ \nabla_{x_2} L = 5 + 1\lambda_1 - 5\lambda_2 - \lambda_3 + \lambda_5 = 0 \\ \nabla_{x_3} L = 2 - 2\lambda_1 - 4\lambda_2 - 2\lambda_3 + \lambda_6 = 0 \\ \lambda \succcurlyeq 0 \\ \lambda_1(2x_1 - x_2 + 2x_3 - 9) = 0 \\ \lambda_2(x_1 + 5x_2 + 4x_3 - 8) = 0 \\ \lambda_3(x_1 + x_2 + 2x_3 - 2) = 0 \\ \lambda_4 x_1 = 0 \\ \lambda_5 x_2 = 0 \\ \lambda_6 x_3 = 0 \\ 2x_1 - x_2 + 2x_3 \leq 9 \\ 3x_1 + 5x_2 + 4x_3 \leq 8 \\ x_1 + x_2 + 2x_3 \leq 2 \\ x_1, x_2, x_3 \geq 0 \end{cases}$$

We have strong duality because KKT conditions are necessary and sufficient. Point $x^* = (1, 1, 0)^T$ is feasible for that.

11 References

1. The best website

2. Convex Optimization, Stephen Boyd

3. The second best website

4. Convex analyses, K. U. Osipenko