Задача 1

Построить двойственную задачу для задачи

$$min_x\sum_{i=1}^{N}||A_ix+b_i||_2+\frac{1}{2}||x-x_0||_2^2$$

где $A_i \in R^{m_i,n}, b_i \in R^{m_i}, x_0 \in R^n$

Воспользуюсь свойством конической двойственности с семинара/лекции. Так как вторая норма негладакая, то возьму всю сложность в конусы. Сделаю замену: $||A_ix+b_i||_2=t_i$

Тогда нашу задачу можно переписать как. $min_{x,t}\sum_{i=1}^{N}t_i+\frac{1}{2}||x-x_0||_2^2, ||A_ix+b_i||_2 \le t_i$.

По определению конуса $K_2 : \begin{pmatrix} A_ix+b_i \\ t_i \end{pmatrix} \in K_2 = \{(x,t)| : ||A_ix+b_i|| \le t_i\}$

Запишу Лагранжиан: $L=\sum_{i=1}^{N}t_i+\frac{1}{2}||x-x_0||_2^2+\sum_{i=1}^{N}(\lambda_i,\bar{\lambda}_i)\begin{pmatrix} A_ix+b_i \\ t_i \end{pmatrix}$

$g=infL$, Теперь возьму производные по x и по t.

$\frac{\partial g}{\partial x}=x^*-x_0+A_1^T\lambda_1+\ldots+A_n^T\lambda_n=0$

Откуда $x^*=x_0-A_1^T\lambda_1-\ldots-A_n^T\lambda_n=x_0-\sum_{i=1}^{N}A_i^T\lambda_i$

$\frac{\partial g}{\partial t_i}=1+\bar{\lambda}_i=0$, откуда $\bar{\lambda}_i=-1$.

Подставлю полученные результаты в g. Тогда
$g=\frac{1}{2}||-\sum_{i=1}^{N}A_i^T\lambda_i||_2^2+\sum_{i=1}^{N}\lambda_i^T(A_ix^*+b_i)=\frac{1}{2}||-\sum_{i=1}^{N}A_i^T\lambda_i||_2^2+\sum_{i=1}^{N}\lambda_i^T(A_i(x_0-\sum_{i=1}^{N}A_i^T\lambda_i)+bi)=\frac{1}{2}||-\sum_{i=1}^{N}A_i^T\lambda_i||_2^2+\sum_{i=1}^{N}\lambda_i^T(A_ix_0+b_i)-\sum_{i=1}^{N}\lambda_i^TA_i\sum_{j=1}^{N}A_j^T\lambda_j$
Замечу, что $\sum_{i=1}^{N}\lambda_i^TA_i\sum_{j=1}^{N}A_j^T\lambda_j=||\sum_{i=1}^{N}A_i^T\lambda_i||_2^2$

Тогда $g=-\frac{1}{2}||\sum_{i=1}^{N}A_i^T\lambda_i||_2^2+\sum_{i=1}^{N}\lambda_i^T(A_ix_0+b_i)$

Так как рассматривалась изначально коническая задача, то надо потребовать, чтобы $\begin{pmatrix} \lambda_i \\ \bar{\lambda}_i \end{pmatrix} \in -K_2^*$. Иначе в силу конического ограничения можем сделать сколько угодно большим слагаемое. Но у нас inf. Тогда двойственную нельзя будет построить.

Тогда $||\lambda_i||_2 \le -\bar{\lambda}_i$,откуда, $||\lambda_i||_2 \le 1$.

Итоговый ответ

$$g=-\frac{1}{2}||\sum_{i=1}^{N}A_i^T\lambda_i||_2^2+\sum_{i=1}^{N}\lambda_i^T(A_ix_0+b_i) \to \max_{\lambda}$$

$$||\lambda_i||_2 \le 1$$

Задача 2

Построить двойственную задачу для задачи: $min_x -\sum_{i=1}^{m}log(b_i-a_i^Tx)$, где $a_i^Tx<b_i, \forall 1 \le i \le m$

Сделаю замену вида: $y_i=b_i-a_i^Tx>0$.

Тогда запишу Лагранжиан.

$L=-\sum_{i=1}^{m}log(y_i)-\sum_{i=1}^{m}\lambda_i(y_i-b_i+a_i^Tx)$

Так как присутствует две переменных, по которым двойственную задачу нужно оптимизировать, то можно оптимизировать по отдельности.(Ссылаюсь на Boyd стр 133 - Optimizing over some variables).

$\frac{\partial g}{\partial x}=\sum_{i=1}^{N}a_i\lambda_i=0$

Далее $g(\lambda)=inf_y(-\sum_{i=1}^{m}log(y_i)-\sum_{i=1}^{m}\lambda_i(y_i-b_i))$

$\frac{\partial g}{\partial y}=-\frac{1}{y_i}-\lambda_i=0$, откуда $y_i=-\frac{1}{\lambda_i}$

Подставлю полученные результаты в изначальную g.

$g(\lambda)=-\sum_{i=1}^{m}(log(-\frac{1}{\lambda_i})-\sum_{i=1}^{m}\lambda_i(-\frac{1}{\lambda_i}-b_i)=\sum_{i=1}^{m}log(-\lambda_i)+m+b^T\lambda$. Буду считать, что $log(0)=0$ (Boyd так делает)

Итого

$$g(\lambda)=\sum_{i=1}^{m}log(-\lambda_i)+m+b^T\lambda \to max, \lambda_i \le 0, \sum_{i=1}^{N}a_i\lambda_i=0$$

Пояснение почему $\lambda_i \le 0$. Пусть это не так и $\lambda_i>0$. Тогда возьму $y_j=0, j \ne i, y_i=t, t>=0$. Тогда $-\lambda_iy_i=-t \to -\infty$. Противоречие.

Замечу, что больше ограничений на λ накладывать не нужно. Для этого рассмотрю $g(\lambda)=inf_y(-\sum_{i=1}^{m}log(y_i)-\sum_{i=1}^{m}\lambda_iy_i)$

Из λ_iy_i знаю, что $\lambda_i \le 0, y_i>=0$, тогда $\inf$ существует и не равен $-\infty$.

Задача 3

Построить двойственную задачу для задачи:

$min -c^Tx+\sum_{i=1}^{m}y_ilogy_i$, где $Px=y, x>=0, \sum_{i=1}^{n}x_i=1$ Где P - это матрица с неотрицательными элементами, а сумма элементов в любом столбце равна 1.

Запишу Лагранжиан.

$L=-c^Tx+\sum_{i=1}^{m}y_ilogy_i+\lambda(\sum_{i=1}^{n}x_i-1)-\mu^Tx+\nu^T(Px-y)$

Буду решать тем же методом что и выше. Буду по отдельности оптимизировать сначала Лагранжиан по x, потом по y.

$\frac{\partial L}{\partial x}=-c+\lambda*1_n-\mu+P^T\nu=0$

Далее $g=inf_y=\sum_{i=1}^{m}y_ilogy_i-\lambda-\nu^Ty=\sum_{i=1}^{m}y_ilogy_i-\lambda-\sum_{i=1}^{m}\nu_iy_i$

$\frac{\partial g}{\partial y}=logy_i+1-\nu_i=0$, откуда $y_i=e^{\nu_i-1}$

Подставлю полученные результаты в изначальную g.

$L=\sum_{i=1}^{m}e^{\nu_i-1}(\nu_i-1)-\lambda-\sum_{i=1}^{m}\nu_ie^{\nu_i-1}=-\sum_{i=1}^{m}e^{\nu_i-1}-\lambda$

Также нужно наложить ограничение на $\nu>=0$. (Так как задача на min, то надо сделать слагаемое $-\nu^Tx$ отрицательным. Но так как $x>=0$, то тогда $\nu>=0$.

Итого двойственная задача имеет вид:

$$g=-\sum_{i=1}^{m}e^{\nu_i-1}-\lambda \to \max_{\nu,\lambda}$$

$$-c+\lambda*1_n-\mu+P^T\nu=0, \nu>=0$$

Задача 5

Оптимальный вписанный эллипсоид может быть получен как решение следующей задачи:

$$min_{B\in S_{++}^n, d\in R^n}-log(detB)$$

$$||Ba_i||_2+a_i^Td \le b_i, 1 \le i \le k$$

Последний шаг, который в условии делать не буду, так как перейду к конической двойственности для $||Ba_i||_2$.

Перейду к эквивалентной задачи оптимизации: $||Ba_i||_2 \le t$

По определению конуса $K_2 : \begin{pmatrix} Ba_i \\ t_i \end{pmatrix} \in K_2 = \{(x,t)| : ||Ba_i|| \le t_i\}$ (Теперь минимизировать кроме $B \in S_{++}^n, d \in R^n$, надо еще по t_i). Запишу Лагранжиан
$L=-log(detB)+\sum_{i=1}^{k}(\lambda_i,\bar{\lambda}_i)\begin{pmatrix} Ba_i \\ t_i \end{pmatrix}+\sum_{i=1}^{k}\nu_i(t_i+a_i^Td-b_i)$

$\frac{\partial L}{\partial d}=\sum_{i=1}^{k}\nu_ia_i=0$

$\frac{\partial L}{\partial t_i}=\bar{\lambda}_i+\nu_i=0$, откуда $\bar{\lambda}_i=-\nu_i$. В силу конического ограничения надо потребовать $\begin{pmatrix} \lambda_i \\ \bar{\lambda}_i \end{pmatrix} \in -K_2^*$. Тогда $||\lambda_i||_2 \le -\bar{\lambda}_i$,откуда, $||\lambda_i||_2 \le \nu_i$.

ТОгда остается $g=inf_{B\in S_{++}^n}=-log(detB)+\sum_{i=1}^{k}\lambda_i^TBa_i-\nu^Tb$

$\frac{\partial log(detB)}{\partial B}=\frac{1}{detB}\frac{\partial detB}{\partial B}=\frac{detB(B^{-T})}{detB}=B^{-T}=B^{-1}$ (В силу симметричности). Тут воспользовался формулой 1.3 отсюда(https://github.com/esokolov/ml-course-hse/blob/master/2020-fall/seminars/sem03-vector-diff.pdf) производной определителя по матрице.

Замечу, что $B=\frac{B+B^T}{2}$

Тогда $\frac{\partial\lambda_iBa_i}{\partial B}=\frac{1}{2}\frac{\partial(\lambda_i^TBa_i+\lambda_i^TB^Ta_i)}{\partial B}=\frac{1}{2}\frac{\partial\lambda_i^TBa_i}{\partial B}+\frac{1}{2}\frac{\partial a_i^TB\lambda_i}{\partial B}$(По свойству следа, так как размерность этой величины равна 1, ничего не изменится$=\frac{1}{2}\lambda_ia_i^T+\frac{1}{2}a_i\lambda_i^T$. Воспользовался свойством 1.5 отсюда (https://github.com/esokolov/ml-course-hse/blob/master/2020-fall/seminars/sem03-vector-diff.pdf)

ТОгда $B^{-1}=\sum_{i=1}^{k}(\frac{1}{2}\lambda_ia_i^T+\frac{1}{2}a_i\lambda_i^T)$

$-log(detB)=log(detB)^{-1}=log(\frac{1}{detB})=log(detB^{-1})=log(det(\sum_{i=1}^{k}(\lambda_ia_i^T+a_i\lambda_i^T)))-log(2^n)$

Итого ответ

$$g=log(det(\sum_{i=1}^{k}(\lambda_ia_i^T+a_i\lambda_i^T)))-nlog(2)-\nu^Tb+\sum_{i=1}^{k}\lambda_i^T(\sum_{i=1}^{k}(\frac{1}{2}\lambda_ia_i^T+\frac{1}{2}a_i\lambda_i^T))^{-1}a_i \to max$$

$$\sum_{i=1}^{k}\nu_ia_i=0, ||\lambda_i||_2 \le \nu_i, \nu>=0$$

$\nu>=0$ из-за неравенства, чтобы сделать слагаемое $\sum_{i=1}^{k}\nu_i(t_i+a_i^Td-b_i)$ отрицательным.

Задача 4

По условию T - непрерывно распределенная на R_+. Тогда ее функция распределения имеет вид $F_T(t)=P(T \le t)$

По условию дана функция выживания $S(t)=P(T \ge t)=exp(-\int_0^t\lambda(\tau)d\tau)$

Тогда функция распределения T имеет вид: $F_T(t)=1-S(t)$

Чтобы составить задачу на максимум правдоподобия, нужно выразить функцию плотности случайной величины T.

$f(t)=F_T'(t)=-S'(t)=\lambda(t)S(t)=\lambda_0(t)e^{w^Tx}e^{-\int_0^t\lambda(\tau)d\tau}=\lambda_0(t)e^{w^Tx}e^{-\int_0^t\lambda_0(\tau)e^{w^Tx}d\tau}$

Составлю функцию правдоподобия $L=\prod_{j=1}^{K}\lambda_0(t_j)e^{w^Tx_j}e^{-\int_0^{t_j}\lambda_0(\tau)e^{w^Tx_j}d\tau}$

Возьму логарифм функции правдоподобия.

$lnL=\sum_{j=1}^{K}\lambda_0(t_j)+\sum_{j=1}^{K}w^Tx_j-\sum_{j=1}^{K}e^{w^Tx_j}\int_0^{t_j}\lambda_0(\tau)d\tau \to max_w$

Константа никак не влияет на w, ее не учитываем. Далее линейная функция выпуклая/вогнутая. Далее экспонента выпуклая функция по w.(Из предыдущей темы) Тогда вычитая из линейной функции выпуклую получаем вогнутую.

Далее по определегию выпуклости: A convex optimization problem is a problem where all of the constraints are convex functions, and the objective is a convex function if minimizing, or a concave function if maximizing. В нашем случае целевая функция - вогнутая и мы ее максимизируем, значит это выпуклая задача оптимизации.

Задача 6

Постройте задачу эквивалентную задачу квадратичного программирования для задачи:

$$\min_{x\in R^2}\max_{1\le i\le k}||y_i-x||_2$$

Сначала в данной задаче находится такой y_i, при котором вторая норма максимальная. Это можно переписать как $||y_i-x||_2 \le R$. И целевая функция. $\min_{x,R}R$.(Тут минимизация по x, так как в искомой задаче берется минимум по x, после того как нашли максимальное расстояние между точками. Но данная задача не является задачей квадратичного программирования. Поэтому перейду к задаче, имеющей такое же решение как и данная.

$$||y_i-x||_2^2 \le R^2$$

$$\min_{x,R}R^2$$

$R \ge 0$ не интересует, так как смотрим на квадрат.

Составлю двойственную задачу для данной.

$L=R^2+\sum_{i=1}^{k}\lambda_i(||y_i-x||_2^2-R^2) \to \min_{x,R}$

$\frac{\partial L}{\partial R}=2R-2R\sum_{i=1}^{k}\lambda_i=0$, откуда $R=0$, или $\sum_{i=1}^{k}\lambda_i=1$

$\frac{\partial L}{\partial x}=\sum_{i=1}^{k}\lambda_iy_i-\sum_{i=1}^{k}\lambda_ix^*=0$ Учитывая, что $\sum_{i=1}^{k}\lambda_i=1$, $x^*=\sum_{i=1}^{k}\lambda_iy_i$

Тогда $dual=g=\sum_{i=1}^{k}\lambda_i||y_i-\sum_{i=1}^{k}\lambda_iy_i||_2^2 \to \max_{\lambda}$

Проверю сильную двойственность.

Подставлю решение в оптимизационную функцию. Так как оптимизируем R^2, то численно оно будет равно ограничению, то есть $||y_i-x^*||_2^2=||y_i-\sum_{i=1}^{k}\lambda_iy_i||_2^2$

Тогда рассмотрю разность полученного значения с двойственной.

$||y_i-\sum_{i=1}^{k}\lambda_iy_i||_2^2-\sum_{i=1}^{k}\lambda_i||y_i-\sum_{i=1}^{k}\lambda_iy_i||_2^2=||y_i-\sum_{i=1}^{k}\lambda_iy_i||_2^2(1-\sum_{i=1}^{k}\lambda_i)=0$. Следовательно сильная двойственность выполняется.

Интерпретация для коэффициентов. Значение по x - афинная оболочка точек y_i.

Так как сильная двойственность выполняется, то подстановка оптимального решения для двойственной задачи в двойственную задачу даст такой же результат как подставление полученного решения для прямой задачи.

Задача 7

$$\max_{x\in R^n}\sum_{j=1}^{m}\pi_jlogp_j^Tx$$

$$x \ge 0, 1^Tx=1$$

Составлю Лагранжиан и решу задачу методом ККТ.

$L=-\sum_{j=1}^{m}\pi_jlogp_j^Tx-\lambda^Tx+\nu(1^Tx-1)$

Условия ККТ для оптимальной точки.

$$-x^* \le 0$$

$$1^Tx^*=1$$

$$\lambda^* \ge 0$$

$$\lambda^*x^*=0$$

$$\frac{\partial L}{\partial x^*}=0$$

$\frac{\partial L}{\partial x^*}=-\frac{\sum_{j=1}^{m}\pi_jp_j^T}{p_j^Tx^*}-\lambda^*+\nu^*1^T=0$, откуда $\frac{\sum_{j=1}^{m}\pi_jp_j^T}{p_j^Tx^*}=\nu^*1^T-\lambda^*$

Замечу, что $\frac{\sum_{j=1}^{m}\pi_jp_j^T}{p_j^Tx^*}x^*=1$ (Сумма вероятностей равна 1).

Тогда можно записать $1=\nu1^Tx^*-\lambda^*x^*$, откуда $\lambda x^*=0$ по условию дополняющей нежесткости.Тогда $1=\nu1^Tx^*, 1=\nu^*1, \nu^*=1$

Теперь могу рассмотреть Лагранжиан покомпонентно с соответсвуюшими условиями для оптимальной точки.

$\frac{\partial L}{\partial x_i^*}=-\frac{\sum_{j=1}^{m}\pi_jp_{ij}}{p_j^Tx^*}-\lambda_i^*+\nu^*=0$

Из условия дополняющей нежесткости $\lambda_i^*x_i^*=0$, следует, что по крайней мере или $\lambda_i^*=0$ или $x_i^*=0$.

Пусть $\lambda_i^*=0$, тогда $\frac{\sum_{j=1}^{m}\pi_jp_{ij}}{p_j^Tx^*}=\nu^*=1$(Из полученного выше)

Пусть $x_i^*=0$, тогда $\frac{\sum_{j=1}^{m}\pi_jp_{ij}}{p_j^Tx^*}=\nu^*-\lambda_i^*=1-\lambda_i^*$, по условию ККТ $\lambda_i^* \ge 0$. Тогда $\frac{\sum_{j=1}^{m}\pi_jp_{ij}}{p_j^Tx^*} \le 1$

Интерпретация. Если $x_i \ge 0$, то математическое ожидание доходности i актива, нормированное на доходность в период для каждого актива i равно 1. Если же $x_i=0$, то данная величина меньше 1, что логично, так как в данном случае капитал в i актив не вкладывался. Значит портфель менее разнообразен.

Запишу двойственную задачу. Так как неудобно дифференцировать искомую функцию по x, введу дополнительную замену. $y=P^Tx$

$L=-\sum_{j=1}^{m}\pi_jlogy_j+\nu^T(y-P^Tx)-\lambda^Tx+\mu(1^Tx-1) \to \min_{y,x}$

$\frac{\partial L}{\partial x}=-\nu^TP^T-\lambda+\mu1^T=0$

$\frac{\partial L}{\partial y_j}=-\frac{\pi_j}{y_j}+\nu_j=0, y_j=\frac{\pi_j}{\nu_j}$

$$dual=g=-\sum_{j=1}^{m}\pi_jlog(\frac{\pi_j}{\nu_j})+\pi-\mu \to max$$

$$-\nu^TP^T-\lambda+\mu1^T=0$$

$$\lambda \ge 0$$

```
In [ ]:
```

```
In [ ]:
```

Тогда $f_\mu(X)=sup_{||Y||_2\le1}tr(X^TY)-\frac{\mu*||Y||_F^2}{2}$

Известно отсюда(https://github.com/esokolov/ml-course-hse/blob/master/2020-fall/seminars/sem03-vector-diff.pdf), что

$$\frac{\partial tr(X^TY)}{\partial Y}=X$$

$$\frac{\partial tr(Y^TY)}{\partial Y}=2Y$$

$\frac{\partial f_\mu(X)}{\partial X}=X-\mu*Y=0, \sigma(X) \le \mu$

$Y=\frac{X}{\mu}, f_\mu(X)=\frac{tr(X^TX)}{\mu}-\frac{tr(X^TX)}{2\mu}=\frac{tr(X^TX)}{2\mu}=\frac{\sigma^2(X)}{2\mu}, \sigma(X) \le \mu$

$f_\mu(X)=tr(X)-\frac{\mu}{2}=\sigma(X)-\frac{\mu}{2}, \sigma(X) \ge \mu$