ESE 605 Homework 4

Xi Cao

Apr. 8, 2022

Problems from Boyd & Vandenberghe:
5.13.

(a) The Lagrange dual is:

$$\begin{aligned} L(x,\nu,\lambda) &= c^T x + \sum \nu_i x_i(1-x_i) + \lambda^T(Ax-b) \\ &= -x^T \mathrm{diag}(\nu)x + (c+\nu+A^T\lambda)^T x - \lambda^T b \end{aligned}$$

To find the lower boundary of L, set:

$$\begin{aligned} \frac{\partial L}{\partial x} &= -2\mathrm{diag}(\nu)^T x + (c+\nu+A^T\lambda) = 0 \\ x &= \frac{c+\nu+A^T\lambda}{2\mathrm{diag}(\nu)} \end{aligned}$$

The dual function is:

$$\begin{aligned} g(\nu,\lambda) &= \inf_x L(x,\nu,\lambda) \\ &= \begin{cases} \frac{(c+\nu+A^T\lambda)^2}{4\mathrm{diag}(\nu)} - \lambda^T b & \nu \preceq 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

So the dual problem is:

$$\begin{array}{ll} \underset{\nu}{\text{maximize}} & \frac{(c+\nu+A^T\lambda)^2}{4\mathrm{diag}(\nu)} - \lambda^T b \\ \text{subject to:} & \nu \preceq 0 \\ & \lambda \succeq 0 \end{array}$$

We can also write this problem as:

$$\begin{array}{ll} \underset{\nu,\lambda}{\text{maximize}} & 1/4\left(\sum_i (c_i+\nu_i+a_i^T\lambda)^2/\nu_i\right) - \lambda^T b \\ \text{subject to:} & \nu \preceq 0 \\ & \lambda \succeq 0 \end{array}$$

Look at $(c_i+\nu_i+a_i^T\lambda)^2/\nu_i$. Since $(c_i+\nu_i+a_i^T\lambda)^2 \geq 0$ and $\nu_i \leq 0$, $(c_i+\nu_i+a_i^T\lambda)^2/\nu_i \leq 0$. If $c_i + a_i^T\lambda \geq 0$, then we can find ν_i such that $c_i + \nu_i + a_i^T\lambda = 0$, its maximum is 0. Otherwise, it reaches its maximum when:

$$\frac{\partial}{\partial \nu_i}(c_i + \nu_i + a_i^T\lambda)^2/\nu_i = 1 - (c_i + a_i^T\lambda)^2/\nu_i = 0$$
$$\nu_i = (c_i + a_i^T\lambda)^2$$
$$\text{maximum} = 4(c_i + a_i^T\lambda)$$

Therefore the maximum value of the dual problem is:

$$\begin{cases} -\lambda^T b & c_i + a_i^T\lambda \geq 0 \\ c_i + a_i^T\lambda - \lambda^T b & c_i + a_i^T\lambda \leq 0 \end{cases}$$

We can write the dual problem as:

$$\begin{array}{ll} \underset{\lambda}{\text{maximize}} & -\lambda^T b + \sum_i \min\{0, c_i + a_i^T\lambda\} \\ \text{subject to:} & \lambda \succeq 0 \end{array}$$

(b) First we want to derive the dual of the LP relaxation.

The Lagrange dual is:

$$\begin{aligned} L(x, \nu_1, \nu_2, \nu_3) &= c^T x + \nu_1^T(Ax - b) + \nu_2^T(-x) + \nu_3^T(x - 1) \\ &= (c + A^T\nu_1 - \nu_2 + \nu_3)^T x + (-b^T\nu_1 - 1^T\nu_3) \end{aligned}$$

The dual function:

$$\begin{aligned} g(\nu_1, \nu_2, \nu_3) &= \inf_x L(x, \nu_1, \nu_2, \nu_3) \\ &= \begin{cases} -b^T\nu_1 - 1^T\nu_3 & c + A^T\nu_1 - \nu_2 + \nu_3 = 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

So the dual problem is:

$$\begin{array}{ll} \underset{\lambda}{\text{maximize}} & -b^T\nu_1 - 1^T\nu_3 \\ \text{subject to:} & c + A^T\nu_1 - \nu_2 + \nu_3 = 0 \\ & \nu_1 \succeq 0,\ \nu_2 \succeq 0,\ \nu_3 \succeq 0 \end{array}$$

If we set:

$$\nu_1 = \lambda$$
$$\nu_{3i} = -(c_i + a_i^T\lambda)$$
$$\nu_2 = 0$$

Then the dual problem of the LP relaxation will give us the same lower bound as the Lagrangian relaxation.

5.17.

We can write $\sup_{a \in \mathcal{P}_i} a^T x$ as a LP:

$$\begin{array}{ll} \underset{\lambda}{\text{maximize}} & a^T x \\ \text{subject to:} & C_i a \preceq d_i \end{array}$$

The Lagrange dual is:

$$\begin{aligned} L(x, \nu) &= a^T x + \nu^T (C_i a - d_i) \\ &= (x + C_i^T \nu)^T a - d_i^T \nu \end{aligned}$$

And the dual function is:

$$\begin{aligned} g(\nu) &= \inf_a L(x, \nu) \\ &= \begin{cases} -d_i^T \nu & x + C_i^T \nu = 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

So the dual problem is:

$$\begin{array}{ll} \underset{\lambda}{\text{maximize}} & -d_i^T \nu \\ \text{subject to:} & -C_i^T \nu = x \\ & \nu \preceq 0 \end{array}$$

Without loss of generality, we can name the optimal solution $-z_i$, so that:

$$\begin{aligned} \sup_{a \in \mathcal{P}_i} a^T x &= d_i^T z_i \\ C_i^T z_i &= x \\ z_i &\succeq 0 \end{aligned}$$

Then the robust LP is equivalent to:

$$\begin{array}{ll} \underset{\lambda}{\text{minimize}} & c^T x \\ \text{subject to:} & d_i^T z_i \leq b_i \\ & C_i^T z_i = x \\ & z_i \succeq 0 \end{array}$$

5.21.

(a) We know that exponential functions are convex, so e^{-x} is a convex function. Also, since $y > 0$, x^2/y is convex over x. Since $x^2 \geq 0$, x^2/y is convex over y. We have verified this is a convex optimization problem.

Since e^{-x} is monotonically decreasing, minimize it is to maximize x. Also, the problem restricts $x^2/y \leq 0$, and from domain restriction, $x^2/y \geq 0$, so $x^2/y = 0$ and $x = 0$. The optimal value is $e^0 = 1$.

(b) The Lagrange dual is:

$$L(x, y, \lambda) = e^{-x} + \lambda(x^2/y)$$

And the dual function is:

$$g(\lambda) = \inf_{x,y} L(x, y, \lambda)$$

To find the infimum of L,

$$\begin{aligned} L_x &= -e^{-x} + 2\lambda x/y \\ L_y &= -\lambda x^2/y^2 \end{aligned}$$

The above two equation cannot both be zero, so L has no extreme points.

If $\lambda < 0$, by setting x large enough and y small enough, the minimum of L can reach $-\infty$. If $\lambda \geq 0$, we know that $e^{-x} \geq 0$ and $\lambda(x^2/y) \geq 0$, so $L(x, y, \lambda) \geq 0$. Also $L(x, y, \lambda)$ can be zero when x goes to infinity and $\lambda = 0$.

So the dual problem is:

$$\begin{aligned} \underset{\lambda}{\text{maximize}} \quad & 0 \\ \text{subject to:} \quad & \lambda \succeq 0 \end{aligned}$$

The optimal value $d^* = 0$, and the optimal duality gap is 1.

(c) From (a) we know the optimal solution of x is $x^* = 0$, then $x^2/y = 0$ and the constraint is not satisfied with strict inequalities. So Slater's condition does not hold for this problem.

(d) When $u = 0$, from (a) we know $p^*(u) = 1$.

When $u > 0$, Slater's condition holds for this perturbed problem, so from (b) we know $p^*(u) = 0$.

When $u < 0$, the problem is infeasible, so $p^*(u) = \infty$.

The global sensitivity inequality does not hold when $u > 0$.

$$0 = p^*(u) < p^*(0) - \lambda^* u = 1 - 0 = 1$$

5.24.

We know that

$$\sup_{z \in Z} \left(\inf_{w \in W} f(w, z) \right) - \epsilon \leq \inf_{w \in W} f(w, z^*)$$

for some $z^* \in Z$ and any $\epsilon > 0$. And the inequation holds even if Z is not feasible. Also,

$$\inf_{w \in W} f(w, z^*) \leq \inf_{w \in W} \left(\sup_{z \in Z} f(w, z) \right)$$

From chain rule we have

$$\sup_{z\in Z}\left(\inf_{w\in W} f(w,z)\right) - \epsilon \leq \inf_{w\in W}\left(\sup_{z\in Z} f(w,z)\right)$$

for any $\epsilon > 0$. Therefore,

$$\sup_{z\in Z}\inf_{w\in W} f(w,z) \leq \inf_{w\in W}\sup_{z\in Z} f(w,z)$$

5.31.

Since f_0 is convex,

$$f_0(x) \geq f_0(x^*) + \nabla f_0\left(x^\star\right)^T\left(x - x^\star\right)$$

$$\begin{aligned}
\nabla f_0\left(x^\star\right)^T\left(x - x^\star\right) &= -\sum_{i=1}^m \lambda_i^* \nabla f_i\left(x^\star\right)\left(x - x^*\right) \\
&= -\sum_{i=1}^m \lambda_i^* f_i\left(x^\star\right) - \sum_{i=1}^m \lambda_i^* \nabla f_i\left(x^\star\right)\left(x - x^*\right) \\
&= -\sum_{i=1}^m \lambda_i^*\left(f_i\left(x^\star\right) + \nabla f_i\left(x^\star\right)\left(x - x^*\right)\right)
\end{aligned}$$

Since $f_i\left(x^\star\right) \leq 0$, $f_i(x) \leq 0$, and f_i is convex,

$$f_i\left(x^\star\right) + \nabla f_i\left(x^\star\right)\left(x - x^*\right) \leq f_i(x) \leq 0$$

Therefore,

$$\nabla f_0\left(x^\star\right)^T\left(x - x^\star\right) \geq 0$$

6.7.

(a)

$$\begin{aligned}
f(x) &= \|Ax - b\|_2^2 + \delta\|x\|_2^2 \\
&= x^T A^T A X - 2b^T A x + b^T b + \delta x^T x \\
&= x^T V \mathbf{diag}(\sigma)^T U^T U \mathbf{diag}(\sigma) V x - 2U\mathbf{diag}(\sigma)V^T b x + b^T b + \delta x^T x \\
&= x^T \mathbf{diag}(\sigma)^T \mathbf{diag}(\sigma) x - 2b^T U \mathbf{diag}(\sigma) V x + b^T b + \delta x^T x \\
&= \sum_{i=1}^r (\sigma_i^2 + \delta)x_i^2 - 2\sigma_i u_i v_i^T b_i x_i + b_i^2
\end{aligned}$$

Take the derivatives,

$$\frac{\partial}{\partial x_i} f(x) = 2(\sigma_i^2 + \delta)x_i - 2b_i\sigma_i u_i v_i^T$$

So the solution is

$$x_i^* = \frac{\sigma_i u_i v_i^T b_i}{\sigma_i^2 + \delta}$$
$$x^* = \sum_{i=1}^{r} \frac{\sigma_i u_i v_i^T b_i}{\sigma_i^2 + \delta}$$

(b) First consider the unconstrained problem:

$$\|Ax - b\|_2^2 = \sum_{i=1}^{r} \sigma_i^2 x_i^2 - 2\sigma_i u_i v_i^T b_i x_i + b_i^2$$
$$x_i^* = \frac{u_i v_i^T b_i}{\sigma_i}$$

If $\|x^*\|_2^2 = \sum_{i=1}^{r} (\frac{u_i v_i^T b_i}{\sigma_i})^2 \leq \gamma$, since rank$(A) = r < min\{m, n\}$, we can find

$$\|x^*\|_2^2 = \sum_{i=1}^{r} (\frac{u_i v_i^T b_i}{\sigma_i})^2 + \sum_{i=r+1}^{n} x_i^2 = r$$

So the solution is equal to the unconstrained solution:

$$x_i^* = \frac{u_i v_i^T b_i}{\sigma_i}$$

If $\|x^*\|_2^2 = \sum_{i=1}^{r} (\frac{u_i v_i^T b_i}{\sigma_i})^2 \geq \gamma$, the optimal solution is similar to the solution in (a):

$$\sum_{i=1}^{n} x_i^2 = \left(\frac{\sigma_i u_i v_i^T b_i}{\sigma_i^2 + \delta} \right)^2 = \gamma$$
$$x_i^* = \frac{\sigma_i u_i v_i^T b_i}{\sigma_i^2 + \delta}$$
$$x^* = \sum_{i=1}^{n} \frac{\sigma_i u_i v_i^T b_i}{\sigma_i^2 + \delta}$$

A6.19.

(a) The optimal objective value is 341.96.

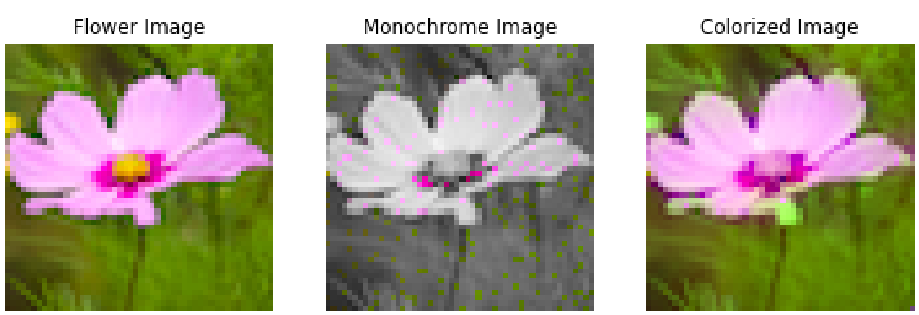
```
import cvxpy as cp
rows, cols, colors = img.shape
given = np.array([R_given.T,G_given.T,B_given.T]).T

variables = []
constraints = []
for i in range(colors):
    U = cp.Variable(shape=(rows, cols))
    variables.append(U)
    constraints.append(U[known_ind] == given[:, :, i][known_ind])
    constraints.append(U<=1)
    constraints.append(U>=0)

constraints.append(0.299*variables[0]+0.587*variables[1]+0.114*variables[2] == M)

prob = cp.Problem(cp.Minimize(cp.tv(*variables)), constraints)
prob.solve(verbose=True, solver=cp.ECOS) #'ECOS', 'SCS', or 'OSQP'.
print("optimal objective value: {}".format(prob.value))
```

```
optimal objective value: 341.96404500840214
```



A7.3.

(a) In this case,

$$
\begin{aligned}
a(\theta) &= \left(\int_{\mathbf{R}_+^n} \exp\left(\theta^T x\right) dx\right)^{-1} \\
&= \left(\frac{\exp\left(\theta^T x\right)}{\prod_{i=1}^n \theta_i} \Big|_{x=0}^{\infty}\right)^{-1} \\
&= \begin{cases} \prod_{i=1}^n -\theta_i & \theta \prec 0 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

For θ to be valid,

$$
\begin{aligned}
\theta &\prec 0 \\
a(\theta) &= \prod_{i=1}^{n} -\theta_i \\
p_\theta(x) &= \prod_{i=1}^{n} (-\theta_i) \exp\left(\theta^T x\right)
\end{aligned}
$$

The associated family of densities is independent exponential distribution.

(b) In this case,

$$
\begin{aligned}
a(\theta) &= (1+\exp(\theta))^{-1} = \frac{1}{1+\exp(\theta)} \\
p_\theta(0) &= \frac{1}{1+\exp(\theta)} \\
p_\theta(1) &= \frac{\exp(\theta)}{1+\exp(\theta)}
\end{aligned}
$$

The associated family of densities is Bernoulli distribution. The valid value of θ is $\theta \in \mathbf{R}$.

(c) If $x \in \mathcal{N}(\mu, \Sigma)$, then it has the density function:

$$
\begin{aligned}
p(x) &= \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{-\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{-\frac{1}{2}}} e^{\left(x^T \Sigma^{-1}\mu - \frac{1}{2}x^T \Sigma^{-1} x - \frac{1}{2}\mu^T \Sigma^{-1}\mu\right)} \\
&= \frac{e^{-\frac{1}{2}\mu^T \Sigma^{-1}\mu}}{(2\pi)^{\frac{n}{2}} |\Sigma|^{-\frac{1}{2}}} e^{\left(x^T \Sigma^{-1}\mu - \frac{1}{2}x^T \Sigma^{-1} x\right)}
\end{aligned}
$$

If we use parameter $(z, Y) = (\Sigma^{-1}\mu, \Sigma^{-1})$, then,

$$
\begin{aligned}
\theta^T c(x) &= x^T \Sigma^{-1}\mu - \frac{1}{2}(x+\mu)^T \Sigma^{-1}(x+\mu) \\
&= z^T x + \mathbf{tr}(Y)(-\frac{1}{2}xx^T)
\end{aligned}
$$

And the density function can be written as:

$$
p_\theta(x) = \frac{e^{-\frac{1}{2}z^T Y^{-1} z}}{(2\pi)^{\frac{n}{2}} |Y|^{\frac{1}{2}}} \exp\left(z^T x + \mathbf{tr}(Y)(-\frac{1}{2}xx^T)\right)
$$

The valid value of z is $\mathbf{R}^n$, and the valid value of Y is $\mathbf{S}^n_{++}$

(d) Consider the case when $\mathcal{D}$ is finite and discrete,

$$\begin{aligned}\log p_\theta(x) &= \log a(\theta) + \theta^T c(x) \\ &= -\log \sum_{x\in\mathcal{D}} \exp\left(\theta^T c(x)\right) + \theta^T c(x)\end{aligned}$$

Since $\theta^T c(x)$ is affine, and we have shown in class that log-sum-exp function is convex, we now have shown that $\log p_\theta(x)$ is concave.

When $\mathcal{D}$ is discrete but infinite or continuous, we can set,

$$\begin{aligned}\mathcal{D} &= \mathcal{D}_1 + \cdots + \mathcal{D}_n + \cdots \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \mathcal{D}_i\end{aligned}$$

and for $x \in \mathcal{D}_i$, $\log p_\theta(x)$ is concave. So the sums of $\log p_\theta(x)$ is concave.

(e) Since the K samples are IID,

$$\begin{aligned}\ell_\theta(x_1, \cdots, x_K) &= \log \prod_{i=1}^{K} p_\theta(x_i) \\ &= \sum_{i=1}^{K} \log p_\theta(x_i) \\ &= \sum_{i=1}^{K} \left(-\log \sum_{x_i\in\mathcal{D}} \exp\left(\theta^T c(x_i)\right) + \theta^T c(x_i) \right)\end{aligned}$$

Therefore,

$$\begin{aligned}\nabla_\theta \ell_\theta(x_1, \cdots, x_K) &= \sum_{i=1}^{K} \left(-\frac{\sum_{x_i\in\mathcal{D}} \exp\left(\theta^T c(x_i)\right) c(x_i)}{\sum_{x_i\in\mathcal{D}} \exp\left(\theta^T c(x_i)\right)} + c(x_i) \right) \\ &= -K\mathbf{E}_\theta c(x) + \sum_{i=1}^{K} c(x_i)\end{aligned}$$

So that

$$(1/K)\nabla_\theta \ell_\theta(x_1, \cdots, x_K) = -\mathbf{E}_\theta c(x) + \frac{1}{K}\sum_{i=1}^{K} c(x_i)$$

A7.7.

```
[143] lam = cp.Variable(p)
      obj = 0
      for k in range(n):
        e = np.zeros(shape = (n, 1))
        e[k] = 1
        obj += 1/m * cp.matrix_frac(e, V @ cp.diag(lam) @ V.T)
      constraints = [cp.sum(lam)==1, cp.min(lam) >= 0]

      prob = cp.Problem(cp.Minimize(obj),constraints)
      prob.solve(solver = cp.CVXOPT)

      print("The optimal value is", prob.value)
```

```
The optimal value is 0.12810192899289358
```

```
low_bnd = prob.value

m_rnd = cp.pos(np.round(m*lam.value,decimals=1))
print(sum(m_rnd.value) == m)

up_bnd = 0
for k in range(n):
  e = np.zeros(shape = (n, 1))
  e[k] = 1
  up_bnd += 1/m * cp.matrix_frac(e, V @ cp.diag(m_rnd/m) @ V.T)


gap = up_bnd - low_bnd
gap_value = gap.value
print ('The gap between the upper and the lower bounds is:', gap_value)
```

```
True
The gap between the upper and the lower bounds is: 4.5609239246735545e-06
```

The gap is really small, which means the relaxed could be a good approximation of discrete A-optimal experiment problem.

A7.21.

(a) The ARX model can be written as:

$$y_{t+1} = \varphi_{t+1}^T \beta + x_{t+1}$$

where,

$$\begin{aligned} \varphi_{t+1} &= [y_t \quad \cdots \quad y_{t-M+1}]^T \quad \in R^M \\ \beta &= [\beta_1 \quad \cdots \quad \beta_M]^T \quad \in R^M \\ x_{t+1} &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

The goal of ARX is to identify β using M consecutive observations. The joint density function of M observations is:

$$f = \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(y_i - \varphi_i^T \beta)^2}{2\sigma^2}\right]$$

The log-likelihood function is:

$$\begin{aligned} L(\beta, \sigma) &= C - \sum_{i=1}^{M} \frac{(y_i - \varphi_i^T \beta)^2}{2\sigma^2} \\ &= C - \frac{1}{2\sigma^2} \sum_{i=1}^{M} (y_i - \varphi_i^T \beta)^2 \end{aligned}$$

where C is constant regardless of the value of β.

Differentiation of the log-likelihood function is:

$$\frac{\partial}{\partial \beta} L(\beta, \sigma) = \frac{1}{2\sigma^2} \sum_{i=1}^{M} 2\varphi_i (y_i - \varphi_i^T \beta)$$

Set the differentiation equals zero, then we get:

$$\sum_{i=1}^{M} \varphi_i (y_i - \varphi_i^T \beta) = 0$$

(b) We want to minimize the error and maximize the likelihood,

$$\text{minimize} \quad \|y - \varphi^T \beta\|_2^2$$

If x is sparse, we can use Lasso Regression as a simple model based on convex optimization. Since the number of equations are more than the number of variables, this is an over-constrained problem:

$$\begin{aligned} \text{minimize} \quad & \|y - \varphi^T \beta\|_2^2 + \lambda \|x\|_1 \\ & = \|y - \varphi^T \beta\|_2^2 + \lambda \|y - \varphi^T \beta\|_1 \end{aligned}$$

(c) We can see the results from part(b) is really close to the true β. In the figure below, they overlapped. According to my limited understanding, this is because we assumed x is sparse, and this prior condition can reduce the interference of noise.

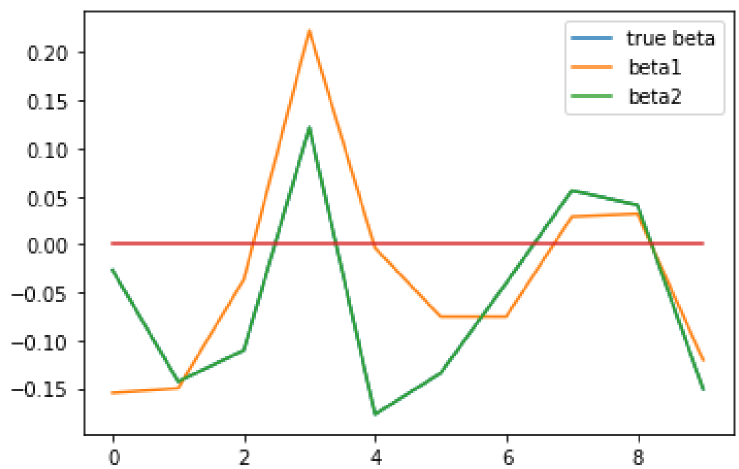
[63]
```
import cvxpy as cp
beta_1 = cp.Variable(shape=(M,1))
obj_1 = 0
for i in range(0,T-M-1):
  y_pre = beta_1.T*y[i+M:i:-1]
  obj_1+=cp.square(y[i+M+1]-y_pre)

prob_1 = cp.Problem(cp.Minimize(obj_1))
prob_1.solve()
```

```
48.65413037425698
```

[75]
```
import cvxpy as cp
beta_2 = cp.Variable(shape=(M,1))
obj_2 = 0
for i in range(0,T-M-1):
  y_pre = beta_2.T*y[i+M:i:-1]
  obj_2+=cp.abs(y[i+M+1]-y_pre)
obj_2 = obj_1+0.1*obj_2
prob_2 = cp.Problem(cp.Minimize(obj_2))
prob_2.solve()
```

```
52.187730968031445
```



A7.37.

(a) Since the N samples are IID, the joint density function is:

$$
\begin{aligned}
f &= \prod_{i=1}^{N} p(x_i; \lambda) \\
&= \prod_{i=1}^{N} \sum_{j=1}^{k} \lambda_j p_j(x)
\end{aligned}
$$

The log-likelihood function is:

$$
L(\lambda) = \sum_{i=1}^{N} \log \left(\sum_{j=1}^{k} \lambda_j p_j(x) \right)
$$

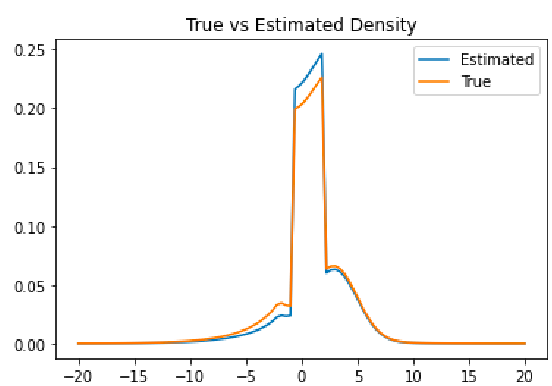
We want to maximize the log-likelihood:

$$
\begin{array}{ll}
\text{minimize} & -\sum_{i=1}^{N} \log \left(\sum_{j=1}^{k} \lambda_j p_j(x) \right) \\
\text{subject to} & 1^T \lambda = 1 \\
& \lambda \succeq 0
\end{array}
$$

```
[ ] lmb = cp.Variable(shape=(k,1))
    obj = -cp.sum(cp.log(lmb.T @ densities.T))
    constraints = [sum(lmb)==1, lmb>=0]
    prob = cp.Problem(cp.Minimize(obj),constraints)
    prob.solve()
```

```
206.88181359033456
```

```
plot_estimated_and_true_density(lmb.value)
```



(b)

A6.5.

(a) The problem is:

$$\begin{aligned} \text{minimize} \quad & \|Ax - b\|_{1.5} \\ & = \left(\sum_{i=1}^{k} |a_i^T x - b_i|^{3/2}\right)^{2/3} \end{aligned}$$

which is equivalent to :

$$\text{minimize} \sum_{i=1}^{k} |a_i^T x - b_i|^{3/2}$$

The gradient of this objective function is:

$$\nabla_x = \sum_{i=1}^{k} (3/2)\text{sign}(a_i^T x - b_i)|a_i^T x - b_i|^{1/2} a_i$$

Set the gradient equals zero, then the optimal condition is:

$$\sum_{i=1}^{k} \text{sign}(a_i^T x - b_i)|a_i^T x - b_i|^{1/2} a_i = 0$$

(b) Combining the methods we used in ℓ_1 and ℓ_2 norms, we can write the problem as:

$$\begin{array}{ll} \underset{\lambda}{\text{minimize}} & 1^T t \\ \text{subject to:} & -s_i \preceq a_i^T x - b_i \preceq s_i \\ & s^{3/2} \preceq t \end{array}$$

Now we have to reform the last constraint to make it a SDP problem.

$$s^{3/2} \preceq t \Leftrightarrow s^2 \preceq t s^{1/2}$$

$$\begin{bmatrix} \sqrt{s_i} & s_i \\ s_i & t \end{bmatrix} \succeq 0$$

To extract the square root function, we introduce y such that $0 \preceq y \preceq \sqrt{s}$, so that we can rewrite the above constraint as:

$$\begin{bmatrix} y_i & s_i \\ s_i & t_i \end{bmatrix} \succeq 0, \quad \begin{bmatrix} s_i & y_i \\ y_i & 1 \end{bmatrix} \succeq 0$$

Now we can formulate the $\ell_{1.5}$ norm approximation problem as an SDP:

$$\begin{array}{ll} \underset{\lambda}{\text{minimize}} & 1^T t \\ \text{subject to:} & -s_i \preceq a_i^T x - b_i \preceq s_i \\ & \begin{bmatrix} y_i & s_i \\ s_i & t_i \end{bmatrix} \succeq 0, \quad \begin{bmatrix} s_i & y_i \\ y_i & 1 \end{bmatrix} \succeq 0 \end{array}$$

(c)

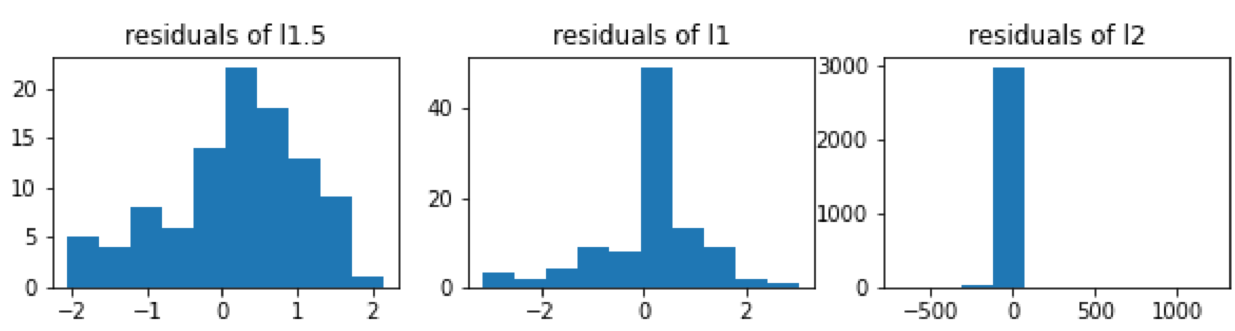
```
# l1.5
x = cp.Variable(shape=(30,1))
s = cp.Variable(shape=(100,1))
obj = cp.mixed_norm(s, 2, 3)
constraints = [s >= cp.abs(A*x-b)]
prob = cp.Problem(cp.Minimize(obj),constraints)
prob.solve()
```

```
4.848858246643649
```

```
[49] # l1
x_1 = cp.Variable(shape=(30,1))
obj_1 = cp.norm(A*x_1-b,1)
prob_1 = cp.Problem(cp.Minimize(obj_1))
prob_1.solve()
```

```
66.15897931542277
```

```
[50] # l2
x_2 = A/b
```



A7.13.

The density of $x = Az + b$ is given by:

$$
\begin{aligned}
p_x(v) &= \frac{1}{|\det A|} p_z(A^{-1}(v-b)) \\
&= \frac{1}{|\det A|} \exp -\phi \left(\left\| A^{-1}(v-b) \right\|_2 \right)
\end{aligned}
$$

$x_1, \cdots, x_N$ are independent samples, and the joint density is:

$$
\left(\frac{1}{|\det A|} \right)^N \prod_{i=1}^{N} \exp(-\phi \left(\left\| A^{-1}(x_i - b) \right\|_2 \right))
$$

The log-likelihood function is:

$$
l(A, b) = -N \log(|\det A|) - \sum_{i=1}^{N} \phi \left(\left\| A^{-1}(x_i - b) \right\|_2 \right)
$$

We notice $l(A, b)$ is not concave. In order to find the maximum value of $l(A, b)$, we use:

$$
B = A^{-1} \qquad c = A^{-1} b
$$

Now the log-likelihood function becomes:

$$
l(B, c) = N \log(\det B) - \sum_{i=1}^{N} \phi \left(\| B(x_i - c) \|_2 \right)
$$

Since $l(B, c)$ is concave, we can find B^*, c^* such that

$$
\begin{aligned}
\frac{\partial}{\partial B} l(B^*, c) = 0 \\
\frac{\partial}{\partial c} l(B, c^*) = 0
\end{aligned}
$$

Therefor, we can get $A^* = (B^*)^{-1}$, $b^* = (B^*)^{-1} c^*$