Dustin Lennon
Math 582 Convex Optimization
Problems from Boyd, Chapter 7

Problem 7.1 Solve the MLE problem when the noise is exponentially distributed with density

$$p(z) = \frac{1}{a} e^{-z/a} \mathbf{1}(z \geq 0)$$

The MLE is given by the following:

$$\begin{aligned}
\prod_{i=1}^{n} p(y_i - a_i' x) &= \prod_{i=1}^{n} (1/a) \exp\left[- \left(\frac{y_i - a_i' x}{a} \right) \right] \mathbf{1}(y_i - a_i' x > 0) \\
&= (1/a)^n \exp\left(- \sum_{i=1}^{n} \frac{y_i - a_i' x}{a} \right), \qquad y \geq Ax
\end{aligned}$$

This means the ML problem can be expressed as

$$\begin{array}{ll}
\inf_x & \sum_{i=1}^n (y_i - a_i' x) \\
\text{s.t.} & y \geq Ax
\end{array}$$

Equivalently,

$$\begin{array}{ll}
\inf_x & 1'(y - Ax) \\
\text{s.t.} & 0 \geq Ax - y
\end{array}$$

This leads to the Lagrangian:

$$\begin{aligned}
L(x, \nu) &= 1'(y - Ax) + \nu'(Ax - y) \\
&= 1'y - \nu' y + (\nu' A - 1' A) x
\end{aligned}$$

And the Lagrange dual,

$$g(\nu) = \begin{cases} (1 - \nu)' y & \text{if } A'(1 - \nu) = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Problem 7.2 Given the linear measurement model, $y = Ax + v$ with uniform noise

$$p(z) = \begin{cases} 1/(2\alpha) & \text{if } |z| \leq \alpha \\ 0 & \text{otherwise} \end{cases}$$

Show that the joint ML estimates of x and α are found by solving the l_∞-norm approximation problem

$$\min \quad \|Ax - y\|_\infty$$

The MLE is computed as the maximum of

$$\prod_{i=1}^{n} \left(\frac{1}{2\alpha} \right) \mathbf{1}(-\alpha \leq a_i' x - y \leq \alpha)$$

This is expressed as,

$$\begin{array}{ll} \max\limits_{x,\alpha} & \left(\frac{1}{2\alpha}\right)^n \\ \text{s.t.} & \|Ax-y\|_\infty < \alpha \end{array}$$

Equivalently, we can solve the minimization problem after a monotonic transformation

$$\begin{array}{ll} \min\limits_{x,\alpha} & \alpha \\ \text{s.t.} & \|Ax-y\|_\infty < \alpha \end{array}$$

And since α is a slack variable, the result follows.

Problem 7.3 Estimate the parameters, a, b in a probit model where v is a zero mean Gaussian variable:

$$y_i = \begin{cases} 1 & a'u_i + b + v_i \le 0 \\ 0 & a'u_i + b + v_i > 0 \end{cases}$$

If $y = 1$, then $a'u_i + b \le v_i$. This event has probability given by $1 - \Phi(a'u_i + b) = \Phi(-a'u_j - b)$ where $\Phi(z) = \int_{-\infty}^{z} \frac{1}{2\pi} \exp\left(-\frac{s^2}{2}\right) ds$. Similarly, for $y = 0$, $a'u_j + b \ge v_j$ has probability given by $\Phi(a'u_j + b)$. We maximize the likelihood equation:

$$\prod_{y_i=0} \Phi(a'u_i + b) \prod_{y_j=1} \Phi(-a'u_j - b)$$

or, equivalently, maximize the log-likelihood equation:

$$\sum_{y_i=0} \log \Phi(a'u_i + b) + \sum_{y_j=1} \log \Phi(-a'u_j - b)$$

Φ is an integral of a log-concave function, hence log-concave. Thus the log likelihood is convex and has the form of a penalty approximation problem.

Problem 7.4a Joint estimation of covariance and mean for a multivariate normal distribution.

Let R be the covariance and a the mean. Define Y and μ to be the respective estimates of R and a:

$$\mu = \tfrac{1}{N} \textstyle\sum_{k=1}^{N} y_k, \qquad Y = \tfrac{1}{N} \textstyle\sum_{k=1}^{N} (y_k - \mu)(y_k - \mu)'$$

First we state the log-likelihood function

$$l(R, a) = -(Nn/2)\log(2\pi) - (N/2)\log\det R - (1/2)\sum_{k=1}^{N} (y_k - a)' R^{-1} (y_k - a)$$

and note that the last term can be rewritten:

$$\begin{aligned} \sum_{k=1}^{N} (y_k - a)' R^{-1} (y_k - a) &= \sum_{k=1}^{N} (y_k - \mu + \mu - a)' R^{-1} (y_k - \mu + \mu - a) \\ &= \sum (y_k - \mu)' R^{-1} (y_k - \mu) + 2 \underbrace{\sum (y_k - \mu)' R^{-1} (\mu - a)}_{=0} + \sum (\mu - a)' R^{-1} (\mu - a) \\ &= \sum \mathbf{tr}(R^{-1}(y_k - \mu)(y_k - \mu)) + N(\mu - a)' R^{-1} (\mu - a) \\ &= N\, \mathbf{tr}(R^{-1} Y) + N(\mu - a)' R^{-1} (\mu - a) \end{aligned}$$

Hence,

$$l(R,a) = -(Nn/2)\log(2\pi) - (N/2)\log\det R - \frac{N}{2}\,\mathbf{tr}(R^{-1}Y) - \frac{N}{2}(\mu - a)'R^{-1}(\mu - a)$$

We maximize the log-likelihood by taking the matrix derivative with respect to R and gradient with respect to a and setting them to zero:

$$\frac{d}{da}(\mu - a)'R^{-1}(\mu - a) = -2R^{-1}(a - \mu)$$
$$\frac{d}{dR}\log\det R = R^{-1}$$
$$\frac{d}{dR}\,\mathbf{tr}(R^{-1}Y) = R^{-1}YR^{-1}$$

where the last two identities are derived by variational methods (see Appendix A in Boyd).

Setting the derivatives equal to zero yields the following ML estimates:

$$a = \mu$$
$$R = Y$$

Problem 7.5a Markov Chain Estimation. Define the transition probability matrix as

$$P_{ij} = \mathbf{prob}\,(y(t+1) = i \mid y(t) = j)$$

where $\sum_{i=1}^{n} P_{ij} = 1$. We write the likelihood function:

$$P(Y_1)\prod_{i=2}^{N} P(y(i) = k_i \mid y(i-1) = k_{i-1})$$

If we denote n_{ij} as the number of transitions from j to i, we can write the above as

$$P(Y_1)\prod_{i,j} P_{ij}^{n_{ij}}$$

With the constraint, this yields a Lagrangian of the log likelihood:

$$L(P, \nu_1, \ldots, \nu_n) = c + \sum_{ij} n_{ij}\log P_{ij} + \sum_j \nu_j\left(1 - \sum_i P_{ij}\right)$$

Taking derivatives, we obtain

$$\frac{\partial L}{\partial P_{ij}} = \frac{n_{ij}}{P_{ij}} - \nu_j = 0$$

And summing over i with the constraint yields $\nu_j = \sum_i n_{ij} \equiv n_j$. Thus the MLE is

$$P_{ij} = \frac{n_{ij}}{n_j}$$

which can be interpreted as the number of observed transitions from j to i divided by the total number of visits to state j.

Problem 7.5b We add the constraint of a known equilibrium distribution, q where $1'q = 1$ and $Pq = q$. This amounts to adding a constraint to the Lagrangian:

$$L(P, \nu, \mu_1, \ldots, \mu_n) = c + \sum_{ij} n_{ij}\log P_{ij} + \sum_j \nu_j\left(1 - \sum_i P_{ij}\right) + \sum_i \mu_i\left(q_i - \sum_j P_{ij}q_j\right)$$

This is convex since both constraints are linear in P_{ij}.

Problem 7.6 Consider a normalized random variable, X, and a shifted and scaled random variable $Y = \frac{X+b}{a}$. We assume that X has density function, $p(x)$. Then the density function of Y can be computed as

$$p_Y(y) = ap_X(ay - b)$$

Hence the likelihood is given by

$$\prod_{i=1}^{n} ap_X(ay_i - b)$$

and the log likelihood is convex since p_X is log-convex.

We compute ML estimates for a, b for the Laplace distribution, $p_x(x) = \exp(-2|x|)$, for which we maximize the log likelihood:

$$\max_{a,b} \quad n \log a - 2 \sum_{i=1}^{n} |ay_i - b|$$

It is equivalent to solve

$$\max_{a} \max_{b} \quad n \log a - 2 \sum_{i=1}^{n} |ay_i - b|$$

The inner maximization yields $b = a\,\mathbf{median}(y_1, \ldots, y_n)$. We can solve for a by taking derivatives. First define $S = \sum |y_i - \mathbf{median}(y_1, \ldots, y_n)|$.

$$\begin{aligned} n \log a - 2aS &\Rightarrow \frac{n}{a} - 2S = 0 \\ &\Rightarrow a = \frac{n}{2S} \\ &\Rightarrow b = \frac{n\,\mathbf{median}(y_1, \ldots, y_n)}{2S} \end{aligned}$$

Problem 7.7

- $X_i \sim \text{Poisson}\,(\mu_i) \quad i \in \{1, \ldots, n\}$.
  These are the types of events.
- p_{ji} probability that the jth device detects an event of type i.
  There are m devices.
- Y_{ji} number of events of type i detected by device j.
  Not directly observed.
- $Y_j = \sum_{i=0}^{n} y_{ji}$ is the number of events detected by device j.
  These are the observations.
- Goal: estimate μ_i via maximum likelihood.

$$\begin{aligned} P\left(Y_{ji} = k\right) &= \sum_{n=0}^{\infty} P\left(Y_{ji} = k | X_i = n\right) P\left(X_i = n\right) \\ &= \sum_{n=0}^{\infty} \left[\binom{n}{k} p_{ji}^{k} (1 - p_{ji})^{n-k}\right] \left[\frac{e^{-\mu_i} \mu_i^n}{n!}\right] \\ &= \underbrace{\sum_{n=0}^{\infty} \frac{e^{-\mu_i}}{e^{-\mu_i p_{ji}} (n-k)!} (\mu_i (1 - p_{ji}))^{n-k}}_{=1} \left[\frac{e^{-\mu_i p_{ji}} (\mu_i p_{ji})^k}{k!}\right] \\ &\sim \text{Poisson}\,(\mu_i p_{ji}) \end{aligned}$$

The sum of independent Poisson random variables with means $\lambda_1, \cdots, \lambda_n$ is a Poisson random variable with mean $\lambda_1 + \cdots + \lambda_n$. Hence,

$$Y_j \sim \text{Poisson}\left(\sum_{i=0}^{n} \mu_i p_{ji}\right)$$

Now form the maximum likelihood estimate,

$$\begin{aligned}
L(\mu|y_i, \ldots, y_m) &= \prod_{j=1}^{m} \frac{\exp\left(-\sum_{i=1}^{n} \mu_i p_{ji}\right)\left(\sum_{i=0}^{n} \mu_i p_{ji}\right)^{y_j}}{y_j!} \\
l(\mu) &= \log\{L(\mu|y_i, \ldots, y_m)\} \\
&= -\sum_{j=1}^{m}\sum_{i=1}^{n} \mu_i p_{ji} + \sum_{j=1}^{m}\left\{y_j \log\left(\sum_{i=1}^{n} \mu_i p_{ji}\right) - \log(y_j!)\right\} \\
&= -\mu' q + \sum_{j=1}^{m}\{y_j \log(\mu' p_j) - \log(y_j!)\}
\end{aligned}$$

The primal problem is then:

$$\min \quad -l(\mu) \qquad \text{such that} \quad -\mu \leq 0$$

This problem is convex, as it is an affine function of μ minus the log of an affine function of μ.

We compute the Lagrangian which is differentiable in μ.

$$\mathcal{L}(\mu, \lambda) = \mu' q - \sum_{j=1}^{m}\{y_j \log(\mu' p_j) - \log(y_j!)\} - \mu' \lambda$$

Problem 7.8 Estimation using sign measurements. Measurements are given by (y_i, a_i, b_i). We estimate x in the model:

$$y_i = \mathbf{sign}(a_i' x + b_i + v_i)$$

where $y_i = \pm 1$ and v_i is a log-concave IID noise term. This problem is very similar to question 7.3. Let $F(x)$ denote the distribution function. Then the MLE is given by

$$\prod_{y_i=1} (1 - F(-a_i' x - b_i) \prod_{y_j=-1} F(-a_j' x - b_j)$$

and the log-likelihood is

$$\sum_{y_i=1} \log((1 - F(-a_i' x - b_i)) + \sum_{y_j=-1} \log(F(-a_j' x - b_j))$$

Since the density is assumed to be log concave, it follows that maximizing the log-likelihood is a convex optimization problem.