Computational Statistics
Chapter 1: Continuous optimization

1. The following data are assumed to be an i.i.d. sample from a Cauchy(θ,1) distribution:

   1.77, -0.23, 2.76, 3.80, 3.47, 56.75, -1.34, 4.24, -2.44, 3.29, 3.71,
   -2.40, 4.53, -0.07, -1.05, -13.87, -2.53, -1.75, 0.27, 43.21

   The density function of the Cauchy(θ,1) distribution is

   $$f(x) = \frac{1}{\pi}\left[(x-\theta)^2 + 1\right]^{-1}. \tag{1}$$

   (a) Draw a box plot and a dot plot of this dataset (use the functions `boxplot()` and `dotchart()`).

   (b) Plot the log-likelihood in the interval $[-10, 10]$. How many modes does it have?

   (c) Program the bisection method in R and apply it to these data with starting points -1 and 1. Use additional runs to explore ways in which the bisection method may fail to find the global maximum.

   (d) Program the Newton's method and apply it to the same data. Study the behavior of the algorithm for different starting points.

   (e) Solve the same problem with the R function `optimize`.

2. The data $(1, 1, 1, 1, 1, 1, 2, 2, 2, 3)$ are assumed to be an i.i.d. sample from a logarithmic distribution,

   $$f(x;\theta) = \frac{\theta^x}{x[-\log(1-\theta)]}, \quad x \in \{1, 2, 3, \ldots\}, \theta \in (0, 1),$$

   with mathematical expectation

   $$\mathbb{E}(X) = \frac{-\theta}{(1-\theta)\log(1-\theta)}.$$

   Estimate θ using

(a) The Newton's method;

(b) The Fisher scoring method.

3. The data `F5_2.txt` from Greene's book "Econometric analysis" is a macroeconomics dataset from the U.S. Department of Commerce. It contains quarterly observations from 1950I to 2000IV of some macroeconomic variables. We want to model the relation between income (variable `realdpi`) and consumption (variable `realcons`) using the following nonlinear model

$$Y = \alpha + \beta Z^\gamma, \tag{2}$$

where Y represents the consumption, Z is the income, and $\theta = (\alpha, \beta, \gamma)^T$ is a vector of parameters.

   (a) Assuming $\gamma = 1$, estimate α and β using linear regression (use the function `lm`). Plot the data with the least squares line, and the residuals. Does the linear model fit the data?

   (b) Considering the nonlinear model, compute the least-squares estimate of θ using the BFGS algorithm implemented in the R function `optim`. Display the results graphically.

   (c) Apply the cyclic coordinate descent algorithm to this problem.

   (d) Apply the Gauss-Newton method to this problem.

   (e) Apply Nelder-Mead algorithm using the R function `optim`, and compare the results obtained by the various optimization startegies.

4. The data `transportation.txt` from Greene's book "Econometric analysis" concern transportation equipment manufacturing in $n = 25$ states of United States. The output variable Y is `ValueAdd` and the two input variables x_1 and x_2 are `Capita` (capital input) and `Labor` (labor input). The stochastic frontier model is

$$\ln Y_i = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + V_i - U_i \tag{3}$$

for $i \in \{1, \ldots, n\}$, where β_0, β_1 and β_2 are coefficients, V_i is an error term assumed to have a normal distribution $\mathcal{N}(0, \sigma_v^2)$ and U_i is a positive inefficiency term having a half-normal distribution $|\mathcal{N}(0, \sigma_u^2)|$ (i.e., the distribution of the absolute value of a normal variable). The log-likelihood function is

$$\ell_y(\theta) = -n \ln \sigma + \frac{n}{2} \log \frac{2}{\pi} - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{\epsilon_i}{\sigma} \right)^2 + \sum_{i=1}^{n} \ln \Phi \left(-\frac{\epsilon_i \lambda}{\sigma} \right), \tag{4}$$

with $\lambda = \sigma_u / \sigma_v$, $\sigma^2 = \sigma_u^2 + \sigma_v^2$, $\theta = (\beta_1, \beta_1, \beta_2, \sigma, \lambda)$, and $\epsilon_i = \ln y_i - (\beta_0 + \beta_1 \ln x_{i1} + \beta_2 \ln x_{i2})$.

(a) Using function `read.table`, store the data as a data frame.

(b) Display a matrix plot of the logarithm of the data (use function `plot`).

(c) Using function `lm`, find the least-squares estimate of $\boldsymbol{\beta}$.

(d) Compute the maximum likelihood (ML) estimate of θ, using function `optim` with the least-squares estimates $\widehat{\boldsymbol{\beta}}_{LS}$ as a starting point.

(e) Plot contours of the log-likelihood by fixing three parameters at their ML values and letting the other two parameters vary (use function `contour`). Verify that the solution found in the previous question is a maximum.

(f) Perform again the optimization with the following starting point: $\theta_0 = (3, \widehat{\boldsymbol{\beta}}_{LS}[2:3], 0.5, 10)$. What do you observe? Check graphically that the solution found is a maximum.

(g) Using random initializations, can you find a better solution than the one obtained when starting from the LS estimate?