Homework 8 solutions

1. Exercise A7.5

Solution.

(a) We write the measurement model as

$$\phi^{-1}(y_i) = a_i^T x + v_i, \quad i = 1, \ldots, m.$$

The function ϕ^{-1} is unknown, but it has derivatives between $1/\beta$ and $1/\alpha$. Therefore the numbers $z_i = \phi^{-1}(y_i)$ and y_i must satisfy the inequalities

$$\frac{y_{i+1} - y_i}{\beta} \leq z_{i+1} - z_i \leq \frac{y_{i+1} - y_i}{\alpha}, \qquad i = 1, \ldots, m-1,$$

if we assume that data points are sorted with y_i in increasing order. Conversely, if z and y satisfy these inequalities, then there exists a nonlinear function ϕ with $y_i = \phi(z_i)$, $i = 1, \ldots, m$, and with derivatives between α and β (for example, a piecewise-linear function that interpolates the points). Therefore, as suggested in the problem statement, we can use $z_1, \ldots, z_m$ as parameters instead of ϕ.

The log-likelihood function is

$$l(z, x) = -\frac{1}{2\sigma^2} \sum_{i=1}^m (z_i - a_i^T x)^2 - m \log(\sigma\sqrt{2\pi}).$$

To find a maximum likelihood estimate of x and z one solves the problem

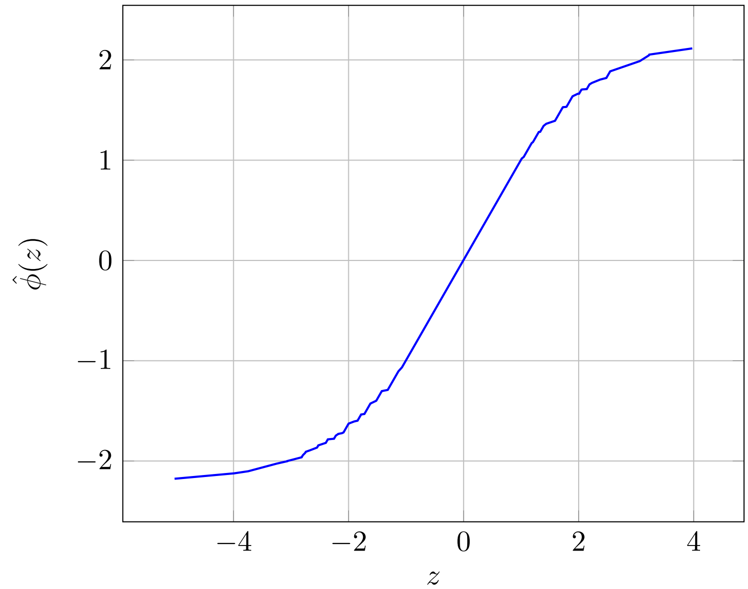
$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m (z_i - a_i^T x)^2 \\ \text{subject to} & (y_{i+1} - y_i)/\beta \leq z_{i+1} - z_i \leq (y_{i+1} - y_i)/\alpha, \quad i = 1, \ldots, m-1. \end{array}$$

This is a quadratic program with variables $z \in \mathbf{R}^m$ and $x \in \mathbf{R}^n$.

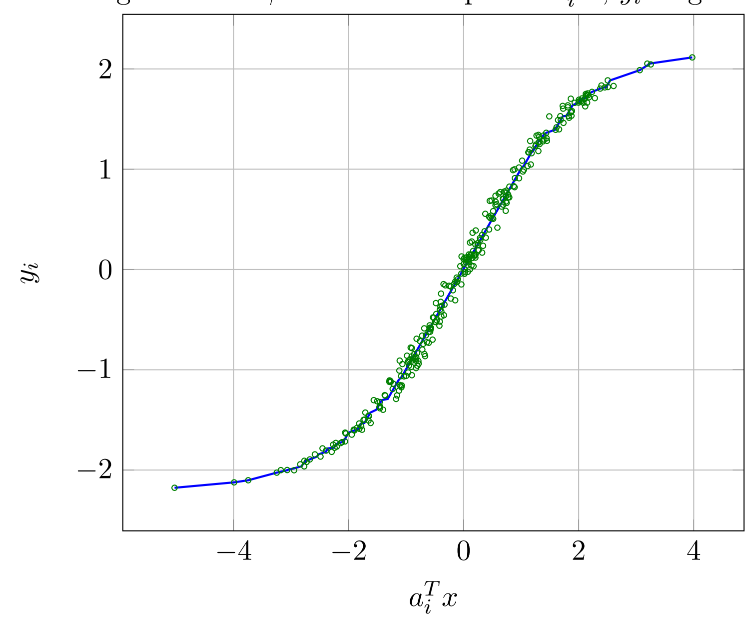
(b) The following MATLAB code solves the problem in the assignment.

```
nonlin_meas_data;
B = [-eye(m-1), zeros(m-1,1)] + [zeros(m-1,1),  eye(m-1)];
cvx_begin
    variables x(n) z(m);
    minimize( norm( z-A*x ) );
    subject to
        (B*y)/beta <= B*z;
        B*z <= (B*y)/alpha;
cvx_end
```

The solution x is $x = (0.4819,\ -0.4657,\ 0.9364,\ 0.9297)$.
The figure shows the estimated function $\hat{\phi}$. This is a piecewise-linear function that satisfies $\hat{\phi}(z_i) = y_i$ for $i = 1, \ldots, m$, and interpolates linearly between the points.



The second figure shows $\hat{\phi}$ and the data points $a_i^T x$, y_i as green circles.



2. Exercise A7.26.

Solution.

(a) We first assume $x \succ 0$, and consider the optimization problem

$$\begin{array}{ll} \text{minimize} & y^T \,\mathbf{diag}(x)^{-1} y \\ \text{subject to} & Ay + c = 0 \end{array}$$

with variable y. Define $X = \mathbf{diag}(x)$. The optimality conditions are

$$2X^{-1}y + A^T u = 0, \qquad Ay + c = 0.$$

From the first equation, $y = -(1/2)XA^T u$. Substituting this in the second equation gives an equation for u:

$$\frac{1}{2}AXA^T u = c.$$

By assumption, c is in the range of AXA^T, so a solution is

$$u = 2(AXA^T)^\dagger c, \qquad y = -XA^T(AXA^T)^\dagger c.$$

Substituting this in the objective gives

$$\begin{aligned} y^T X^{-1} y &= c^T(AXA^T)^\dagger AXX^{-1}XA^T(AXA^T)^\dagger c \\ &= c^T(AXA^T)^\dagger(AXA^T)(AXA^T)^\dagger c \\ &= c^T(AXA^T)^\dagger c. \end{aligned}$$

This expression is still correct if x has zero components. If $x_k = 0$, then necessarily $y_k = 0$ at the optimum, because otherwise $h(x_k, y_k) = \infty$. The nonzero components of y_k are the solution of

$$\begin{array}{ll} \text{minimize} & \sum_{k\in I} y_k^2/x_k \\ \text{subject to} & \sum_{k\in I} y_k a_k + c = 0, \end{array}$$

where $I = \{k \mid x_k > 0\}$. This is the same problem as considered above, but with A replaced by its submatrix of columns indexed by I. We then find that for the optimal y,

$$\begin{aligned} \sum_k h(x_k, y_k) &= \sum_{k\in I} \frac{y_k^2}{x_k} \\ &= c^T(\sum_{k\in I} x_k a_k a_k^T)^\dagger c \\ &= c^T(A\,\mathbf{diag}(x)A^T)^\dagger c. \end{aligned}$$

We conclude that if we optimize over y in problem (33) of the assignment, the problem reduces to (32), so the two problems are equivalent.

(b) We first assume that $y_k \neq 0$ for all k. Consider the problem

$$\begin{array}{ll} \text{minimize} & \sum_{k=1}^n y_k^2/x_k \\ \text{subject to} & \mathbf{1}^T x = 1, \end{array}$$

with implicit constraint $x \succ 0$. The optimality conditions are

$$x \succ 0, \qquad \mathbf{1}^T x = 1, \qquad \frac{y_k^2}{x_k^2} = \nu, \quad k = 1, \ldots, n.$$

From the second equation, $x_k = |y_k|/\sqrt{\nu}$. Substituting this in $\mathbf{1}^T x = 1$ shows that $\sqrt{\nu} = \|y\|_1$. Therefore $x_k = |y_k|/\|y\|_1$. Making this substitition in the cost function (33) gives

$$\sum_{k=1}^{n} \frac{y_k^2}{x_k} = \|y\|_1^2,$$

so the problem reduces to (34).

This conclusion remains valid if y has zero elements. Suppose that at an optimal solution for (33), y has a zero component $y_k = 0$ and $x_k > 0$. Since $y_k^2/x_k = 0$ for all nonnegative values of x_k, we can set $x_k = 0$ without changing the objective value. Then $\mathbf{1}^T x < 1$, so we can increase a component x_j for which $y_j \neq 0$ and this decreases the cost function. (At least one component of y is nonzero, because $A^T y = -c \neq 0$.) Therefore at the optimum, $x_k = 0$ whenever $y_k = 0$, so the expression $x_k = |y_k|/\|y\|_1$ is still correct.

3. Exercise A8.1.

Solution.

(a) The ellipsoid $\mathcal{E} = \{Q^{1/2}y \mid \|y\|_2 \le 1\}$ is contained in C if and only if

$$\|Q^{1/2}a_i\|_2 = \sup_{\|y\|_2 \le 1} |a_i^T Q^{1/2} y| \le 1, \quad i = 1, \ldots, p.$$

(b) The dual function is

$$\begin{aligned} g(\lambda) &= \inf_{Q \succ 0} L(Q, \lambda) \\ &= \inf_{Q \succ 0} \left(\log\det Q^{-1} + \sum_{i=1}^{p} \lambda_i (a_i^T Q a_i - 1) \right) \\ &= \inf_{Q \succ 0} \left(\log\det Q^{-1} + \mathbf{tr}\left((\sum_{i=1}^{p} \lambda_i a_i a_i^T) Q \right) - \sum_{i=1}^{p} \lambda_i \right). \end{aligned}$$

We now use the following fact:

$$\inf_{X \succ 0} \left(\log\det X^{-1} + \mathbf{tr}(XY) \right) = \begin{cases} \log\det Y + n & Y \succ 0 \\ -\infty & \text{otherwise.} \end{cases}$$

The value for $Y \succ 0$ follows by setting the gradient of $\log\det X^{-1} + \mathbf{tr}(XY)$ to zero. This gives $-X^{-1} + Y = 0$, so the minimizer is $X = Y^{-1}$ if $Y \succ 0$. If $Y \not\succ 0$, there exists a nonzero a with $a^T Y a \le 0$. Choosing $X = I + taa^T$ gives $\det X = 1 + t\|a\|_2^2$ and

$$\log\det X^{-1} + \mathbf{tr}(XY) = -\log(1 + ta^T a) + \mathbf{tr}\, Y + ta^T Y a.$$

If $a^T Y a \le 0$ this goes to $-\infty$ as $t \to \infty$.

We conclude that the dual function is

$$g(\lambda) = \begin{cases} \log\det \sum_{i=1}^{p} (\lambda_i a_i a_i^T) - \sum_{i=1}^{p} \lambda_i + n & \text{if } \sum_{i=1}^{p} (\lambda_i a_i a_i^T) \succ 0 \\ -\infty & \text{otherwise.} \end{cases}$$

The resulting dual problem is

$$\begin{array}{ll} \text{maximize} & \log\det \sum_{i=1}^{p}(\lambda_i a_i a_i^T) - \sum_{i=1}^{p} \lambda_i + n \\ \text{subject to} & \lambda \succeq 0. \end{array}$$

(c) The KKT conditions are:

- *Q is primal feasible:* $Q \succ 0$ and $a_i^T Q a_i \leq 1$ for $i = 1, \ldots, p$.
- *Dual multipliers are nonnegativity:* $\lambda \succeq 0$.
- *Complementary slackness:* $\lambda_i(1 - a_i^T Q a_i) = 0$ for $i = 1, \ldots, p$.
- *Gradient of Lagrangian is zero:*

$$Q^{-1} = \sum_{i=1}^{p} \lambda_i a_i a_i^T. \tag{1}$$

The complementary slackness condition implies that $a_i^T Q a_i = 1$ if $\lambda_i > 0$. Now suppose Q and λ are primal and dual optimal. If we take the inner product of the two sides of the equation (1) with Q, we get

$$n = \sum_{i=1}^{p} \lambda_i \, \mathbf{tr}(Q a_i a_i^T) = \sum_{i=1}^{p} \lambda_i a_i^T Q a_i = \sum_{i=1}^{p} \lambda_i.$$

The last step follows from the complementary slackness conditions. Finally, we note, again using (1), that

$$x^T Q^{-1} x = \sum_{i=1}^{p} \lambda_i (a_i^T x)^2 \leq \sum_{i=1}^{p} \lambda_i = n$$

if $x \in C$, *i.e.*, if $|a_i^T x| \leq 1$ for $i = 1, \ldots, p$.

4. *Maximum likelihood estimation from quantized measurements.*