L. Vandenberghe 4/17/2021
ECE133A

Homework 2 solutions

1. *Exercise A1.7.*

$$\begin{aligned}
J &= \frac{1}{n}\|c_1\mathbf{1} + c_2 a - b\|^2 \\
&= \frac{1}{n}\|c_2(a - m_a\mathbf{1}) - (b - m_b\mathbf{1})\|^2 \\
&= \frac{1}{n}(c_2^2\|a - m_a\mathbf{1}\|^2 + \|b - m_b\mathbf{1}\|^2 - 2c_2(a - m_a\mathbf{1})^T(b - m_b\mathbf{1})) \\
&= c_2^2 s_a^2 + s_b^2 - 2c_2\rho s_a s_b \\
&= \rho^2 s_b^2 + s_b^2 - 2\rho^2 s_b^2 \\
&= (1 - \rho^2)s_b^2.
\end{aligned}$$

On line 2, we use $c_1 = m_b - c_2 m_a$. On line 4, we use the definitions of s_a, s_b, and ρ. On line 5, we use $c_2 = \rho s_b/s_a$.

2. *Exercise A1.8.*

(a) We first expand the square in the numerator of J:

$$\begin{aligned}
J &= \frac{(c_1\mathbf{1} + c_2 a - b)^T(c_1\mathbf{1} + c_2 a - b)}{n(1 + c_2^2)} \\
&= \frac{c_1^2 n + 2c_1\mathbf{1}^T(c_2 a - b) + \|c_2 a - b\|^2}{n(1 + c_2^2)} \\
&= \frac{c_1^2 + 2c_1(c_2 m_a - m_b) + \|c_2 a - b\|^2}{(1 + c_2^2)}.
\end{aligned}$$

The derivative with respect to c_1 is

$$\frac{2(c_1 + c_2 m_a - m_b)}{(1 + c_2^2)}.$$

Setting this to zero gives $c_1 = m_b - m_a c_2$. The orthogonal distance regression line passes through the point of averages (m_a, m_b) in the scatter plot: $m_b = c_1 + c_2 m_a$.

(b) The numerator in the expression for J is

$$\begin{aligned}
&\|c_2(a - m_a\mathbf{1}) - (b - m_b)\mathbf{1}\|^2 \\
&\quad = c_2^2\|a - m_a\mathbf{1}\|^2 + \|b - m_b\mathbf{1}\|^2 - 2c_2(a - m_a\mathbf{1})^T(b - m_b\mathbf{1}) \\
&\quad = n\left(c_2^2 s_a^2 + s_b^2 - 2\rho c_2 s_a s_b\right).
\end{aligned}$$

Setting the derivative of J with respect to c_2 to zero gives an equation

$$\frac{2s_a^2 c_2 - 2\rho s_a s_b}{1+c_2^2} - \frac{2c_2(s_a^2 c_2^2 + s_b^2 - 2\rho s_a s_b c_2)}{(1+c_2^2)^2} = 0.$$

After simplifications this reduces to the quadratic equation in the problem statement.

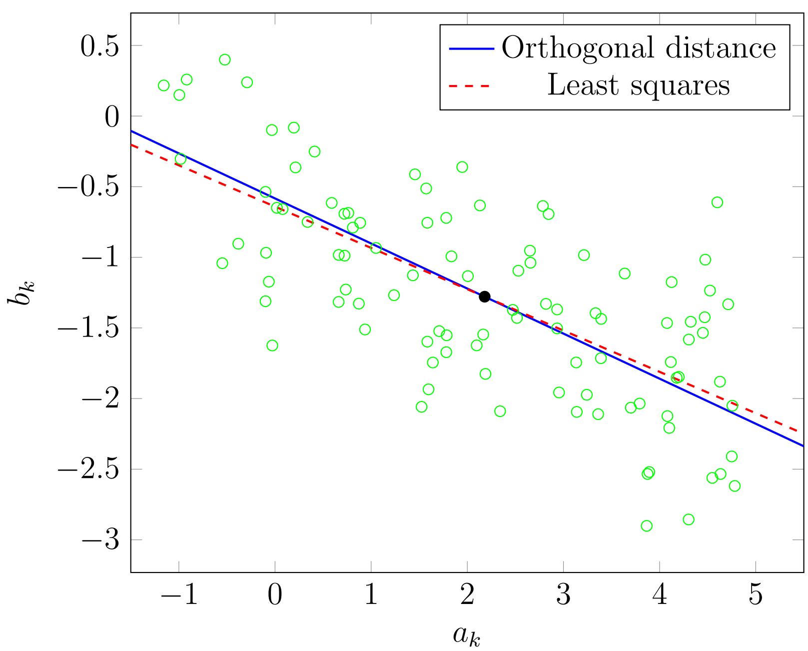
To see that the root with the same sign as ρ is the correct one, note that J approaches a limit s_a^2 as $c_2 \to \infty$ and $c_2 \to -\infty$. Also the derivative has two roots, one positive and one negative. One of these roots must correspond to a maximum of J and the other to a minimum. If $\rho > 0$ the derivative of J at $c_2 = 0$ is negative. Therefore the positive root is a minimum. If $\rho < 0$ the derivative of J at $c_2 = 0$ is positive, and the negative root is a minimum.

This choice also makes intuitive sense, because we expect the optimal slope to have the same sign as ρ, as it does for the (least squares) regression line (see, for example, the scatter plots on page 40 of lecture 2).

(c) The table gives the two solutions.

	c_1	c_2
Least squares	-0.64117	-0.29243
Orthogonal distance	-0.58323	-0.31900

The solid line in the figure is the orthogonal distance regression fit. The dashed line is the least squares fit. The black dot indicates the point of averages (m_a, m_b).



3. *Exercise A2.4.* Define $B = I + A$. This is a nonnegative matrix, with the same off-diagonal elements as A and positive elements on the diagonal.

The diagonal elements of B^2 are

$$(B^2)_{ii} = \sum_{k=1}^{n} B_{ik}B_{ki} = B_{ii}^2 + \sum_{k\neq i} B_{ik}B_{ki}.$$

This is always positive because B_{ii} is positive.

Consider the off-diagonal element $(B^2)_{ij}$ with $i \neq j$:

$$(B^2)_{ij} = B_{ii}B_{ij} + B_{ij}B_{jj} + \sum_{\substack{k\neq i\\ k\neq j}} B_{ik}B_{kj}.$$

Since B_{ii} and B_{jj} are positive, the first two terms are nonzero if and only if $B_{ij} > 0$, *i.e.*, there exists an arc $j \to i$. The sum on the right-hand side is positive if and only if there is at least one k with $B_{ik} > 0$ and $B_{kj} > 0$, *i.e.*, there exists a directed path $j \to k \to i$ of length two. We can summarize this by saying that $(B^2)_{ij}$ is positive if and only if there is a directed path of length two or less from vertex j to vertex i.

In a similar way, one proves that the diagonal elements of B^m are positive, and that an off-diagonal element $(B^m)_{ij}$ is positive if and only if there is a directed path of length m or less from vertex j to vertex i. This can be seen by induction. Suppose the statement is correct for B^{m-1}. The diagonal element $(B^m)_{ii}$ is

$$(B^m)_{ii} = (B^{m-1})_{ii}B_{ii} + \sum_{k\neq i}(B^{m-1})_{ik}B_{ki}.$$

This is always positive because $B_{ii} > 0$ and $(B^{m-1})_{ii} > 0$. Consider $i \neq j$. The element $(B^m)_{ij}$ is

$$(B^m)_{ij} = \sum_{k=1}^{n}(B^{m-1})_{ik}B_{kj} = (B^{m-1})_{ii}B_{ij} + (B^{m-1})_{ij}B_{jj} + \sum_{\substack{k\neq i\\ k\neq j}}(B^{m-1})_{ik}B_{kj}.$$

The first term is positive if there is an arc $j \to i$. The second term is positive if $(B^m)_{ij} > 0$, *i.e.*, there is a directed path of length $m-1$ or less from j to i. The term $(B^{m-1})_{ik}B_{kj}$ in the sum is positive if there is an arc $j \to k$ and a directed path of length $m-1$ or less from k to i. Together, these form a directed path of length m or less from j to i that starts with the arc $j \to i$. All these cases can be summarized by saying that $(B^m)_{ij} > 0$ if and only if there is a directed path of length m or less from j to i.

In a strongly connected graph, there is a directed path from every vertex to every other vertex. Since there are only n vertices, it is sufficient to consider paths of length $n-1$ or less. In a strongly connected graph there is a path of length $n-1$ or less from every vertex to every other vertex. Equivalently, $B^{n-1} = (I + A)^{n-1}$ has positive elements.

4. *Exercise A2.8.* We first compute the n matrix-vector products

$$z_1 = Bx_1, \quad z_2 = Bx_2, \quad \ldots, \quad z_n = Bx_n.$$

This takes $n(2n^2) = 2n^3$ flops. We then compute

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} A_{11}z_1 + A_{12}z_2 + \cdots + A_{1n}z_n \\ \vdots \\ A_{n1}z_1 + A_{n2}z_2 + \cdots + A_{nn}z_n \end{bmatrix},$$

where we partitioned y in subvectors $y_1, \ldots, y_n$ of size n. For each y_k this involves n scalar-vector products $A_{ki}z_i$ and $n-1$ vector additions of size n. This takes $2n^2$ operations, so the total for computing $y_1, \ldots, y_n$ is $2n^3$.

The total flop count is $4n^3$, an order less than for a general product of this dimension (which takes $2(n^2)^2 = 2n^4$ flops).

To derive the complexity, we can also note that

$$\begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = B \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} A^T,$$

so y can be computed using two matrix-matrix products of size $n \times n$, *i.e.*, in $4n^3$ flops.

5. (a) If we compute the product $A_iA_{i+1}\cdots A_j$ as

$$A_iA_{i+1}\cdots A_j = (A_i \cdots A_k)(A_{k+1}\cdots A_j),$$

the number of flops is

$$c_{ik} + c_{k+1,j} + 2n_{i-1}n_kn_j.$$

The first term is the cost of computing $A_i \cdots A_k$, the second term is the cost of computing $A_{k+1}\cdots A_j$, and the third term is for the product of these two matrices, which have dimensions $n_{i-1} \times n_k$ and $n_k \times n_j$, respectively. The best choice of k is the one that minimizes $c_{ik} + c_{k+1,j} + 2n_{i-1}n_kn_j$.

(b) We find

$$\begin{array}{llll} c_{11} = 0 & c_{12} = 10^{10} & c_{13} = 1.2\ 10^{10} & c_{14} = 1.21\ 10^9 \\ & c_{22} = 0 & c_{23} = 10^{11} & c_{24} = 1.20\ 10^9 \\ & & c_{33} = 0 & c_{34} = 2\ 10^8 \\ & & & c_{44} = 0. \end{array}$$

So the flop counts for $A_1A_2A_3A_4$ and $A_1A_2A_3$ are

$$c_{14} = 1.21\ 10^9 \text{ flops}, \qquad c_{13} = 1.20\ 10^{10} \text{ flops},$$

respectively. We note that $A_1A_2A_3A_4$ is cheaper to compute than $A_1A_2A_3$. This is explained by the fact that the optimal order is different. For the first three matrices, the optimal order is $(A_1A_2)A_3$. For the four matrices, the optimal order is $A_1(A_2(A_3A_4))$ and this does not require computing $A_1A_2A_3$.