CS205 Homework #2 Solutions

Problem 1

[Heath 3.29, page 152] Let $\mathbf{v}$ be a nonzero n-vector. The hyperplane normal to $\mathbf{v}$ is the (n-1)-dimensional subspace of all vectors $\mathbf{z}$ such that $\mathbf{v^T z} = \mathbf{0}$. A *reflector* is a linear transformation $\mathbf{R}$ such that $\mathbf{Rx} = -\mathbf{x}$ if $\mathbf{x}$ is a scalar multiple of $\mathbf{v}$, and $\mathbf{Rx} = \mathbf{x}$ if $\mathbf{v^T x} = \mathbf{0}$. Thus, the hyperplane acts as a mirror: for any vector, its component within the hyperplane is invariant, whereas its component orthogonal to the hyperplane is reversed.

1. Show that $\mathbf{R} = \mathbf{2P} - \mathbf{I}$, where $\mathbf{P}$ is the orthogonal projector onto the hyperplane normal to $\mathbf{v}$. Draw a picture to illustrate this result

2. Show that $\mathbf{R}$ is symmetric and orthogonal

3. Show that the Householder transformation

$$\mathbf{H} = \mathbf{I} - \mathbf{2}\frac{\mathbf{vv^T}}{\mathbf{v^T v}},$$

is a reflector

4. Show that for any two vectors $\mathbf{s}$ and $\mathbf{t}$ such that $\mathbf{s} \neq \mathbf{t}$ and $\|\mathbf{s}\|_\mathbf{2} = \|\mathbf{t}\|_\mathbf{2}$, there is a reflector $\mathbf{R}$ such that $\mathbf{Rs} = \mathbf{t}$

Solution

1. We can obtain the reflection $\mathbf{Rx}$ of a vector $\mathbf{x}$ with respect to a hyperplane through the origin by adding to $\mathbf{x}$ twice the vector from $\mathbf{x}$ to $\mathbf{Px}$, where $\mathbf{Px}$ is the projection of $\mathbf{x}$ onto the same hyperplane (see figure 1). Thus

$$\mathbf{Rx} = \mathbf{x} + \mathbf{2}\left(\mathbf{Px} - \mathbf{x}\right) = \left(\mathbf{2Px} - \mathbf{x}\right) = \left(\mathbf{2P} - \mathbf{I}\right)\mathbf{x}$$

Since this has to hold for all $\mathbf{x}$ we have $\mathbf{R} = \mathbf{2P} - \mathbf{I}$.

An alternative way to derive the same result is to observe that the projection $\mathbf{Px}$ lies halfway between $\mathbf{x}$ and its reflection $\mathbf{Rx}$. Therefore

$$\frac{1}{2}\left(\mathbf{x} + \mathbf{Rx}\right) = \mathbf{Px} \Rightarrow \mathbf{Rx} = \left(\mathbf{2Px} - \mathbf{x}\right) = \left(\mathbf{2P} - \mathbf{I}\right)\mathbf{x}$$

which leads to the same result.

2. A reflection with respect to a hyperplane through the origin does not change the magnitude of the reflected vector (see figure 1). Therefore we have

$$\|\mathbf{Rx}\| = \|\mathbf{x}\| \Rightarrow \mathbf{x^T R^T R x} = \mathbf{x^T x} \Rightarrow \mathbf{x^T}(\mathbf{R^T R} - \mathbf{I})\mathbf{x} = \mathbf{0}$$

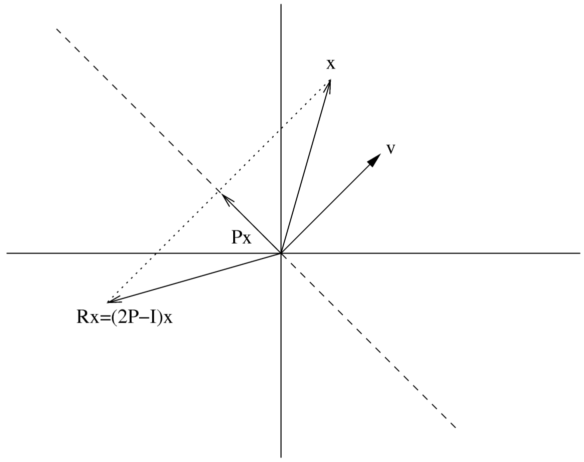

Figure 1: Reflector

for any vector $\mathbf{x}$. If we could show that $\mathbf{x^T}(\mathbf{R^T R} - \mathbf{I})\mathbf{x} = \mathbf{0}$ implies $\mathbf{R^T R} - \mathbf{I} = \mathbf{0}$ we would have proven the orthogonality of $\mathbf{R}$. Furthermore, since reflecting a vector twice just gives the original vector we have $\mathbf{R^2} = \mathbf{I}$. Therefore we would have

$$\mathbf{R^T R} = \mathbf{I} \Rightarrow \mathbf{R^T R^2} = \mathbf{R} \Rightarrow \mathbf{R^T} = \mathbf{R}$$

which shows that $\mathbf{R}$ is symmetric.

In order to show that $\mathbf{R^T R} - \mathbf{I} = \mathbf{0}$ it suffices to show that for a symmetric matrix $\mathbf{C}$, $\mathbf{x^T C x} = \mathbf{0}$ for all $\mathbf{x}$ implies $\mathbf{C} = \mathbf{0}$ (since $\mathbf{R^T R} - \mathbf{I}$ is symmetric). To show that, we note that $\mathbf{e_i^T C e_j} = \mathbf{C_{ij}}$ where $\mathbf{e_k}$ is the k-th column of the identity matrix. We have $\mathbf{e_i^T C e_i} = \mathbf{C_{ii}} = \mathbf{0}$ for any i and furthermore

$$0 = (\mathbf{e_i} + \mathbf{e_j})^T \, \mathbf{C} \, (\mathbf{e_i} + \mathbf{e_j}) = \mathbf{C_{ii}} + \mathbf{C_{jj}} + \mathbf{C_{ij}} + \mathbf{C_{ji}} = \mathbf{2C_{ij}} = \mathbf{2C_{ji}} \Rightarrow \mathbf{C} = \mathbf{0}$$

3. The Householder matrix reflects all vectors in the direction of $\mathbf{v}$

$$\mathbf{H}(\alpha\mathbf{v}) = \left(\mathbf{I} - \mathbf{2}\frac{\mathbf{v v^T}}{\mathbf{v^T v}}\right)(\alpha\mathbf{v}) = \alpha\mathbf{v} - \mathbf{2}\alpha\frac{\mathbf{v}(\mathbf{v^T v})}{\mathbf{v^T v}} = \alpha(\mathbf{v} - \mathbf{2v}) = -(\alpha\mathbf{v})$$

and leaves all vectors $\mathbf{x}$ with $\mathbf{v^T x} = \mathbf{0}$ invariant

$$\mathbf{H x} = \left(\mathbf{I} - \mathbf{2}\frac{\mathbf{v v^T}}{\mathbf{v^T v}}\right)\mathbf{x} = \mathbf{x} - \mathbf{2}\frac{\mathbf{v}(\mathbf{v^T x})}{\mathbf{v^T v}} = \mathbf{x}$$

therefore, $\mathbf{H}$ is a reflector about the hyperplane $\{x : \mathbf{v^T x} = \mathbf{0}\}$.

4. Any two vectors $\mathbf{s}$ and $\mathbf{t}$ are reflections of each other with respect to the hyperplane normal to the vector $\mathbf{s} - \mathbf{t}$ that passes from the midpoint of $\mathbf{s}$ and $\mathbf{t}$. When $\|\mathbf{s}\|_\mathbf{2} = \|\mathbf{t}\|_\mathbf{2}$

that hyperplane passes through the origin and can be written as $\{x : (\mathbf{s}-\mathbf{t})^{\mathbf{T}}\mathbf{x} = \mathbf{0}\}$. Therefore the Householder transform

$$\mathbf{H} = \mathbf{I} - \mathbf{2}\frac{(\mathbf{s}-\mathbf{t})(\mathbf{s}-\mathbf{t})^{\mathbf{T}}}{(\mathbf{s}-\mathbf{t})^{\mathbf{T}}(\mathbf{s}-\mathbf{t})}$$

is the reflection that maps $\mathbf{s}$ to $\mathbf{t}$ and vice versa.

To show that formally, we have

$$\begin{aligned}
\mathbf{Hs} &= \mathbf{s} - \mathbf{2}\frac{(\mathbf{s}-\mathbf{t})(\mathbf{s}-\mathbf{t})^{\mathbf{T}}\mathbf{s}}{(\mathbf{s}-\mathbf{t})^{\mathbf{T}}(\mathbf{s}-\mathbf{t})} = \frac{\mathbf{s}(\mathbf{s}-\mathbf{t})^{\mathbf{T}}(\mathbf{s}-\mathbf{t}) - \mathbf{2}(\mathbf{s}-\mathbf{t})(\mathbf{s}-\mathbf{t})^{\mathbf{T}}\mathbf{s}}{(\mathbf{s}-\mathbf{t})^{\mathbf{T}}(\mathbf{s}-\mathbf{t})} \\
&= \frac{\mathbf{ss^Ts} - \mathbf{2st^Ts} + \mathbf{st^Tt} - \mathbf{2ss^Ts} + \mathbf{2st^Ts} + \mathbf{2ts^Ts} - \mathbf{2tt^Ts}}{(\mathbf{s}-\mathbf{t})^{\mathbf{T}}(\mathbf{s}-\mathbf{t})} \\
&= \frac{2\mathbf{ts^Ts} - \mathbf{2tt^Ts}}{(\mathbf{s}-\mathbf{t})^{\mathbf{T}}(\mathbf{s}-\mathbf{t})} \quad \begin{bmatrix} \mathbf{ss^Ts} + \mathbf{st^Tt} - \mathbf{2ss^Ts} = \mathbf{0} \\ -2\mathbf{st^Ts} + \mathbf{2st^Ts} = \mathbf{0} \end{bmatrix} \\
&= \frac{\mathbf{ts^Ts} - \mathbf{2tt^Ts} + \mathbf{tt^Tt}}{(\mathbf{s}-\mathbf{t})^{\mathbf{T}}(\mathbf{s}-\mathbf{t})} = \frac{\mathbf{t}(\mathbf{s}-\mathbf{t})^{\mathbf{T}}(\mathbf{s}-\mathbf{t})}{(\mathbf{s}-\mathbf{t})^{\mathbf{T}}(\mathbf{s}-\mathbf{t})} = \mathbf{t}
\end{aligned}$$

Problem 2

Let $\mathbf{A}$ be a rectangular $m \times n$ matrix with full column rank and $m > n$. Consider the $\mathbf{QR}$ decomposition of $\mathbf{A}$.

1. Show that $\mathbf{P_0} = \mathbf{I} - \mathbf{QQ^T}$ is the projection matrix onto the nullspace of $\mathbf{A^T}$
2. Show that for every $\mathbf{x}$ we have $\|\mathbf{Ax} - \mathbf{b}\|_\mathbf{2}^\mathbf{2} = \|\mathbf{A}(\mathbf{x} - \mathbf{x_0})\|_\mathbf{2}^\mathbf{2} + \|\mathbf{Ax_0} - \mathbf{b}\|_\mathbf{2}^\mathbf{2}$ where $\mathbf{x_0}$ is the least squares solution of $\mathbf{Ax} = \mathbf{b}$
3. Show that the minimum value for the 2-norm of the residual is attained when $\mathbf{x}$ is equal to the least squares solution and that this minimum value is equal to $\|\mathbf{P_0b}\|_\mathbf{2}$

Solution

1. We know that the nullspace of $\mathbf{A^T}$ and the column space of $\mathbf{A}$ are the normal complements of each other. Therefore, any vector $\mathbf{x}$ can be written as $\mathbf{x} = \mathbf{x_1} + \mathbf{x_2}$ where $\mathbf{x_1}$ is in the nullspace of $\mathbf{A^T}$ and $\mathbf{x_2}$ is in the column space of $\mathbf{A}$.

   For $\mathbf{x_1}$ this means that $\mathbf{A^Tx_1} = \mathbf{0}$. Since $\mathbf{A}$ is full rank, the $\mathbf{QR}$ decomposition is defined and

   $$\mathbf{A^Tx_1} = \mathbf{0} \Rightarrow \mathbf{R^TQ^Tx_1} = \mathbf{0} \Rightarrow \mathbf{Q^Tx_1} = \mathbf{0}$$

   since $\mathbf{R}$ is nonsingular. On the other hand, $\mathbf{x_2}$ belongs to the column space of $\mathbf{A}$, therefore it can be written as $\mathbf{x_2} = \mathbf{Ay} = \mathbf{QRy}$ where $\mathbf{y} \in \mathbf{R^n}$.

Thus, the action of $\mathbf{P_0}$ on $\mathbf{x}$ amounts to

$$\begin{aligned}\mathbf{P_0x} &= \left(\mathbf{I}-\mathbf{QQ^T}\right)(\mathbf{x_1}+\mathbf{x_2}) = \mathbf{x_1}+\mathbf{x_2}-\mathbf{QQ^Tx_1}-\mathbf{QQ^TQRy} \\ &= \mathbf{x_1}+\mathbf{x_2}-\mathbf{QRy} = \mathbf{x_1}\end{aligned}$$

Therefore, $\mathbf{P_0}$ is the projection matrix onto the nullspace of $\mathbf{A^T}$ (and $\mathbf{QQ^T}$ is the projection matrix onto the column space of $\mathbf{A}$).

2. We have

$$\begin{aligned}\|\mathbf{Ax}-\mathbf{b}\|_\mathbf{2}^\mathbf{2} &= \|\mathbf{A}(\mathbf{x}-\mathbf{x_0})+(\mathbf{Ax_0}-\mathbf{b})\|_\mathbf{2}^\mathbf{2} \\ &= \left[\mathbf{A}(\mathbf{x}-\mathbf{x_0})+(\mathbf{Ax_0}-\mathbf{b})\right]^T\left[\mathbf{A}(\mathbf{x}-\mathbf{x_0})+(\mathbf{Ax_0}-\mathbf{b})\right] \\ &= \|\mathbf{A}(\mathbf{x}-\mathbf{x_0})\|_\mathbf{2}^\mathbf{2}+\|\mathbf{Ax_0}-\mathbf{b}\|_\mathbf{2}^\mathbf{2}+\mathbf{2}(\mathbf{x}-\mathbf{x_0})^\mathbf{T}\mathbf{A^T}(\mathbf{Ax_0}-\mathbf{b}) \\ &= \|\mathbf{A}(\mathbf{x}-\mathbf{x_0})\|_\mathbf{2}^\mathbf{2}+\|\mathbf{Ax_0}-\mathbf{b}\|_\mathbf{2}^\mathbf{2}+\mathbf{2}(\mathbf{x}-\mathbf{x_0})^\mathbf{T}(\mathbf{A^TAx_0}-\mathbf{A^Tb}) \\ &= \|\mathbf{A}(\mathbf{x}-\mathbf{x_0})\|_\mathbf{2}^\mathbf{2}+\|\mathbf{Ax_0}-\mathbf{b}\|_\mathbf{2}^\mathbf{2} \qquad [\mathbf{A^TAx_0}-\mathbf{A^Tb}=\mathbf{0}]\end{aligned}$$

3. From the equation above, we have that the minimum value for $\|\mathbf{Ax}-\mathbf{b}\|_\mathbf{2}$ is attained for $\mathbf{x}=\mathbf{x_0}$, since the term $\|\mathbf{Ax_0}-\mathbf{b}\|_\mathbf{2}^\mathbf{2}$ does not depend on $\mathbf{x}$. The least squares solution is given as

$$\mathbf{Rx_0}=\mathbf{Q^Tb} \Rightarrow \mathbf{x_0}=\mathbf{R^{-1}Q^Tb}$$

Therefore the minimum value for the residual $\|\mathbf{Ax}-\mathbf{b}\|_\mathbf{2}$ is

$$\begin{aligned}\|\mathbf{Ax_0}-\mathbf{b}\|_\mathbf{2} &= \|\mathbf{QRR^{-1}Q^Tb}-\mathbf{b}\|_\mathbf{2} = \|\mathbf{QQ^Tb}-\mathbf{b}\|_\mathbf{2} = \left\|(\mathbf{QQ^T}-\mathbf{I})\mathbf{b}\right\|_\mathbf{2} \\ &= \left\|(\mathbf{I}-\mathbf{QQ^T})\mathbf{b}\right\|_2 = \|\mathbf{P_0b}\|\end{aligned}$$

Intuitively, this means that the least squares solution annihilates the component of the residual in the column space of $\mathbf{A}$ and the minimum value for the residual is exactly the component of $\mathbf{b}$ that is *not* contained in the column space of $\mathbf{A}$.

Problem 3

State whether the following classes of matrices are positive (semi-)definite, negative (semi-)definite, indefinite, or whether their definiteness cannot be determined in general

1. Orthogonal matrices
2. Matrices of the form $\mathbf{A^TA}$ where $\mathbf{A}$ is a rectangular matrix
3. Projection matrices
4. Matrices of the form $\mathbf{I}-\mathbf{P}$ where $\mathbf{P}$ is a projection matrix
5. Householder matrices

6. Upper triangular matrices with positive diagonal elements

7. A diagonally dominant matrix with positive elements on the diagonal. A matrix is called diagonally dominant if $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$ and $|a_{ii}| > \sum_{j \neq i} |a_{ji}|$.

Solution

1. Any diagonal matrix with values $+1$ or -1 on the diagonal is orthogonal. Nevertheless it can be positive definite (if it equals $\mathbf{I}$), negative definite (if it equals $-\mathbf{I}$) or indefinite (in any other case). Thus the definiteness of orthogonal matrices cannot be determined for the general case.

2. We have $\mathbf{x^T}(\mathbf{A^T A})\mathbf{x} = \|\mathbf{Ax}\|_\mathbf{2}^\mathbf{2} \geq \mathbf{0}$. Thus a matrix of the form $\mathbf{A^T A}$ is always positive semidefinite. In addition, if $\mathbf{A}$ is full rank, then $\mathbf{A^T A}$ is positive definite (since $\mathbf{Ax} = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}$).

3. Let V be the vector subspace that a projection matrix $\mathbf{P}$ projects onto, and $V^\perp$ its normal complement. Let $\mathbf{x} = \mathbf{x_1} + \mathbf{x_2}$ be an arbitrary vector, where $\mathbf{x_1}$ is the component of $\mathbf{x}$ in V and $\mathbf{x_2}$ its component in $V^\perp$. Therefore

$$\mathbf{x^T P x} = (\mathbf{x_1} + \mathbf{x_2})^\mathbf{T} \mathbf{P} (\mathbf{x_1} + \mathbf{x_2}) = \mathbf{x_1^T P x_1} = \|\mathbf{x_1}\|_\mathbf{2}^\mathbf{2} \geq \mathbf{0}$$

   where we used the fact that $\mathbf{P}$ is symmetric and $\mathbf{P x_2} = \mathbf{0}$. Therefore a projection matrix is always positive semi-definite.

4. The matrix $\mathbf{I} - \mathbf{P}$ is the projection onto the *normal complement* of the space $\mathbf{P}$ projects onto. Therefore it is a projection matrix itself and thus positive semidefinite.

5. Given the Householder matrix

$$\mathbf{H} = \mathbf{I} - \mathbf{2}\frac{\mathbf{v v^T}}{\mathbf{v^T v}}$$

   we have $\mathbf{v^T H v} = \mathbf{v^T}(-\mathbf{v}) = -\|\mathbf{v}\|_\mathbf{2}^\mathbf{2} < \mathbf{0}$ where if $\mathbf{w}$ is a nonzero vector that is orthogonal to $\mathbf{v}$ (such a vector always exists in 2 or more dimensions) then $\mathbf{w^T H w} = \mathbf{w^T w} = \|\mathbf{w}\|_\mathbf{2}^\mathbf{2} > \mathbf{0}$. Therefore a Householder matrix is always indefinite (in the special 1D case the matrix reduces to the single number -1, being negative definite)

6. $\mathbf{I}$ is an example of such a matrix that is positive definite. The matrix

$$\begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix}$$

   is an example of an indefinite matrix. However, we can conclude that such a matrix can never be negative definite, because $\mathbf{e_i^T A e_i} = \mathbf{A_{ii}} > \mathbf{0}$, where $\mathbf{e_i}$ is the i-th column of the identity matrix.

7.

$$\begin{aligned}
\mathbf{x^T A x} &= \sum_{i,j} \mathbf{A}_{ij} x_i x_j = \sum_i \mathbf{A}_{ii} x_i^2 + \sum_{i,j\neq i} \mathbf{A}_{ij} x_i x_j \\
&\geq \sum_i |\mathbf{A}_{ii}||x_i|^2 - \sum_{i,j\neq i} |\mathbf{A}_{ij}||x_i||x_j| \\
&= \frac{1}{2}\sum_i (|\mathbf{A}_{ii}| + |A_{ii}|)\,|x_i|^2 - \sum_{i,j\neq i} |\mathbf{A}_{ij}||x_i||x_j| \\
&> \frac{1}{2}\sum_{i,j\neq i} (|\mathbf{A}_{ij}| + |\mathbf{A}_{ji}|)\,|x_i|^2 - \sum_{i,j\neq i} |\mathbf{A}_{ij}||x_i||x_j| \\
&= \sum_{i,j\neq i} |\mathbf{A}_{ij}| \left(\frac{1}{2}|x_i|^2 + \frac{1}{2}|x_j|^2 - |x_i||x_j|\right) \\
&= \frac{1}{2}\sum_{i,j\neq i} |\mathbf{A}_{ij}| \left(|x_i| - |x_j|\right)^2 \geq 0
\end{aligned}$$

Problem 4

[Heath 3.12 page 150]

1. Let $\mathbf{A}$ be a $n \times n$ matrix. Show that any two of the following conditions imply the other.
   (a) $\mathbf{A^T} = \mathbf{A}$
   (b) $\mathbf{A^T A} = \mathbf{I}$
   (c) $\mathbf{A^2} = \mathbf{I}$
2. Give a specific example, other than the identity matrix $\mathbf{I}$ or a permutation of it, of a 3×3 matrix that has all three of these properties.
3. Name a nontrivial class of matrices that have all three of these properties.

Solution

1.

$$\left\{ \begin{array}{c} \mathbf{A^T} = \mathbf{A} \\ \mathbf{A^T A} = \mathbf{I} \end{array} \right\} \Rightarrow \mathbf{A^2} = \mathbf{I} \quad \text{[By substitution]}$$

$$\left\{ \begin{array}{c} \mathbf{A^T} = \mathbf{A} \\ \mathbf{A^2} = \mathbf{I} \end{array} \right\} \Rightarrow \mathbf{A^T A} = \mathbf{I} \quad \text{[By substitution]}$$

$$\left\{ \begin{array}{c} \mathbf{A^T A} = \mathbf{I} \\ \mathbf{A^2} = \mathbf{I} \end{array} \right\} \Rightarrow \left\{ \begin{array}{c} \mathbf{A^{-1}} = \mathbf{A^T} \\ \mathbf{A^{-1}} = \mathbf{A} \end{array} \right\} \Rightarrow \mathbf{A^T} = \mathbf{A}$$

2.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \begin{bmatrix} 1/3 & -2/3 & -2/3 \\ -2/3 & 1/3 & -2/3 \\ -2/3 & -2/3 & 1/3 \end{bmatrix}$$

3. Reflection matrices (e.g. Householder matrices, see problem 1).

Problem 5

[Heath 3.16 page 150] Consider the vector $\mathbf{a}$ as an $n \times 1$ matrix.

1. Write out its QR factorization, showing the matrices $\mathbf{Q}$ and $\mathbf{R}$ explicitly.
2. What is the solution to the linear least squared problem $\mathbf{ax} \cong \mathbf{b}$, where $\mathbf{b}$ is a given n-vector?

Solution

1. By simple application of the algorithm, we have

$$\mathbf{Q} = \frac{\mathbf{1}}{\|\mathbf{a}\|_{\mathbf{2}}}\mathbf{a} \qquad \mathbf{R} = [\|\mathbf{a}\|_{\mathbf{2}}]$$

2. The least squares solution is given by the equation

$$\mathbf{R}x = \mathbf{Q^T b} \Rightarrow \|\mathbf{a}\|_{\mathbf{2}} \cdot x = \frac{\mathbf{1}}{\|\mathbf{a}\|_{\mathbf{2}}}\mathbf{a^T b} \Rightarrow x = \frac{\mathbf{a^T b}}{\mathbf{a^T a}}$$