Homework 4

Symmetric positive definiteness, Cholesky factorization

1. (Heath 2.37) Suppose that the symmetric $(n+1) \times (n+1)$ matrix

$$B = \begin{pmatrix} \alpha & \mathbf{a}^T \\ \mathbf{a} & A \end{pmatrix}$$

is positive definite.

(a) Show that the scalar α must be positive and the $n \times n$ matrix A must be positive definite.

(b) What is the Cholesky factorization of B in terms of α, $\mathbf{a}$, and the Cholesky factorization of $A - \frac{1}{\alpha}\mathbf{a}\mathbf{a}^T$?

Solution:

(a) From the definition of positive definite, $\mathbf{v}^T B \mathbf{v} > 0$ for all $\mathbf{v} \in \mathbb{R}^{n+1}, \mathbf{v} \neq \mathbf{0}$. Let $\mathbf{v} = \begin{pmatrix} x \\ \mathbf{y} \end{pmatrix}$, where x is a scalar and $\mathbf{y}$ is an n-vector. Then we have

$$\begin{aligned} \mathbf{v}^T B \mathbf{v} &= \begin{pmatrix} x & \mathbf{y}^T \end{pmatrix} \begin{pmatrix} \alpha & \mathbf{a}^T \\ \mathbf{a} & A \end{pmatrix} \begin{pmatrix} x \\ \mathbf{y} \end{pmatrix} \\ &= x^2\alpha + \mathbf{y}^T\mathbf{a}x + x\mathbf{a}^T\mathbf{y} + \mathbf{y}^T A\mathbf{y} \\ &= \alpha x^2 + 2x\mathbf{a}^T\mathbf{y} + \mathbf{y}^T A \mathbf{y} > 0 \end{aligned}$$

If we choose $\mathbf{y} = \mathbf{0}$ then the only remaining nonzero term is αx^2, so α must be positive. If instead $x = 0$, then we are left with $\mathbf{y}^T A\mathbf{y}$, and for this to be positive A must be symmetric positive definite by definition.

(b) Let the Cholesky factorization of $A - \frac{1}{\alpha}\mathbf{a}\mathbf{a}^\mathbf{T}$ be $\mathbf{L}\mathbf{L}^\mathbf{T}$. then

$$B = \begin{pmatrix} \sqrt{\alpha} & 0 \\ \frac{\mathbf{a}}{\sqrt{\alpha}} & \mathbf{L} \end{pmatrix} \begin{pmatrix} \sqrt{\alpha} & \frac{\mathbf{a}^\mathbf{T}}{\sqrt{\alpha}} \\ 0 & \mathbf{L}^\mathbf{T} \end{pmatrix} = \begin{pmatrix} \alpha & \mathbf{a}^T \\ \mathbf{a} & A - \frac{1}{\alpha}\mathbf{a}\mathbf{a}^\mathbf{T} + \frac{1}{\alpha}\mathbf{a}\mathbf{a}^\mathbf{T} \end{pmatrix} = \begin{pmatrix} \alpha & \mathbf{a}^T \\ \mathbf{a} & A \end{pmatrix}$$

Sparse matrices

2. *LU factorization and fill-in.* Consider a sparse matrix generated in Matlab (or Octave) as follows

```
n = 100;
A = diag(rand(n,1));
A(1,:) = rand(1,n);
A(:,1) = rand(1,n);
```

(a) Use the command `spy(A)` to visualize the sparsity pattern of A.

(b) Compute the LU factorization of the matrix using the following function `my_lu`. Run `spy(L)` and `spy(U)` . Are L and U also sparse? Now try the built-in Matlab command *lu* on A. Are the L and U generated by the Matlab command sparse?

```
function [L,U] = my_lu(A)
n = size(A,1);
L = zeros(size(A));
A2 = A;
for k = 1:n
    if A2(k,k) == 0
        'Encountered 0 pivot. Stopping'
        return
    end
    L(k,k) = 1;
    for i = k+1:n
        L(i,k) = A2(i,k)/A2(k,k);
    end
    for i = k+1:n
        for j = k+1:n
            A2(i,j) = A2(i,j) - L(i,k)*A2(k,j);
        end
    end
end
U = triu(A2);
end
```

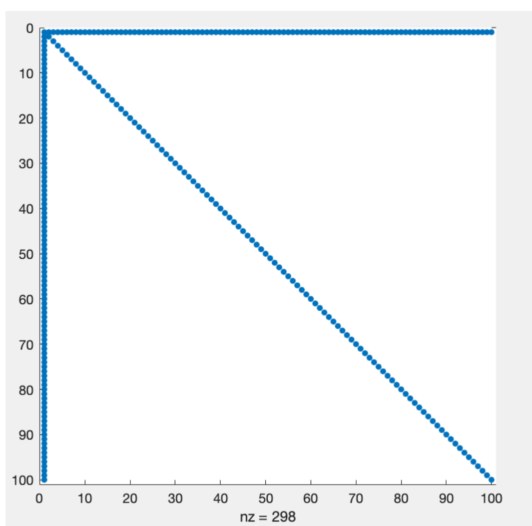
(c) A *tridiagonal* matrix is a matrix that has non-zeros only on its main diagonal and its first sub- and super-diagonals. It is an example of a *banded* matrix. Consider the tridiagonal matrix B given by the following Matlab commands:

```
n = 100;
B = diag(10*ones(n,1)) + diag(3*ones(n-1,1),1) + diag(2*ones(n-1,1),-1);
```

Run `my_lu` and Matlab's `lu` on this matrix. What are the sparsity patterns of L and U using the two different commands?

Solution:

(a) You were not asked to include the figure here, but `spy(A)` produces the figure below.



(b) No, L and U are not sparse. `spy` shows that L is a dense lower triangular matrix and U is a dense upper triangular matrix. This naive implementation of LU factorization produces significant fill-in in the factors L and U, which is bad since we'd prefer to exploit the sparsity for lower storage and computational costs. On the other hand, the L and U generate by Matlab are indeed sparse matrices. Matlab's `lu` is finding a row permutation of the matrix that A that helps to preseve the sparsity in the L and U. This is desirable since the sparse L and U will require significantly fewer operations in the forward and backward solves, in addition to requiring less storage if a sparse matrix representation is used internally.

(c) In this case, the L and U factors produced by both `my_lu` and Matlab's `lu` are sparse, with L having only a nonzero main diagonal and first subdiagonal, and U a nonzero main diagonal and first super diagonal. For banded matrices, the naive LU implementation leads to an L and U each with half the bandwith of A.

The code that produces these results is as follows:

```
n = 100;
A = diag(rand(n,1));
A(1,:) = rand(1,n);
A(:,1) = rand(1,n);

spy(A);
[L,U] = my_lu(A);
spy(L)
spy(U)

[Lm,Um] = lu(A);
spy(Lm)
spy(Um

n = 100;
B = diag(10*ones(n,1)) + diag(3*ones(n-1,1),1) + diag(2*ones(n-1,1),-1);

spy(B);
[L,U] = my_lu(B);
spy(L)
spy(U)

[Lm,Um] = lu(B);
spy(Lm)
spy(Um)
```

Matrix and vector norms

3. Let $\mathbf{x} \in \mathbb{R}^n$. Two vector norms, $||\mathbf{x}||_a$ and $||\mathbf{x}||_b$, are *equivalent* if $\exists\, c_1, d_1 \in \mathbb{R}$ such that

$$c_1||\mathbf{x}||_b \leq ||\mathbf{x}||_a \leq d_1||\mathbf{x}||_b.$$

Matrix norm equivalence is defined analogously to vector norm equivalence, i.e., $||\cdot||_a$ and $||\cdot||_b$ are equivalent if $\exists\, c_2, d_2$ s.t. $c_2||A||_b \leq ||A||_a \leq d_2||A||_b$.

(a) Let $\mathbf{x} \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$. For each of the following, verify the inequality and give an example of a non-zero vector or matrix for which the bound is achieved (showing that the bound is tight):

i. $||\mathbf{x}||_\infty \leq ||\mathbf{x}||_2$

ii. $||\mathbf{x}||_2 \leq \sqrt{n}||\mathbf{x}||_\infty$

iii. $||A||_\infty \leq \sqrt{n}||A||_2$

iv. $||A||_2 \leq \sqrt{n}||A||_\infty$

This shows that $||\cdot||_\infty$ and $||\cdot||_2$ are equivalent, and that their induced matrix norms are equivalent.

(b) Prove that the equivalence of two vector norms implies the equivalence of their induced matrix norms.

Solution:

(a) i. The bound holds because

$$||\mathbf{x}||_\infty = \max_{1\le i\le n} |x_i| = \left(\max_{1\le i\le n} |x_i|^2\right)^{\frac{1}{2}} \le \left(\sum_{i=1}^n |x_i|^2\right)^{\frac{1}{2}} = ||\mathbf{x}||_2.$$

An example vector for which $||\mathbf{x}||_\infty = ||\mathbf{x}||_2$ is given by $\mathbf{x} = (1, 0, \ldots, 0)^T$.

ii. The bound holds because

$$||\mathbf{x}||_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{\frac{1}{2}} \le \left(\sum_{i=1}^n \max_{1\le j\le n} |x_j|^2\right)^{\frac{1}{2}} = \left(n \max_{1\le j\le n} |x_j|^2\right)^{\frac{1}{2}} = \sqrt{n}||\mathbf{x}||_\infty$$

An example vector for which $||\mathbf{x}||_2 = \sqrt{n}||\mathbf{x}||_\infty$ is given by $\mathbf{x} = (1, 1, \ldots, 1)^T$.

iii. Recall that

$$||A||_\infty = \max_{||\mathbf{x}||_\infty = 1} ||A\mathbf{x}||_\infty,$$

and let

$$\mathbf{y} = \operatorname*{argmax}_{||\mathbf{x}||_\infty = 1} ||A\mathbf{x}||_\infty.$$

Then

$$||A||_\infty = ||A\mathbf{y}||_\infty \le ||A\mathbf{y}||_2 = \frac{||A\mathbf{y}||_2}{||\mathbf{y}||_\infty} \le \frac{||A\mathbf{y}||_2}{\frac{1}{\sqrt{n}}||\mathbf{y}||_2} = \sqrt{n}\frac{||A\mathbf{y}||_2}{||\mathbf{y}||_2} \le \sqrt{n}||A||_2,$$

where the first inequality follows from i, the second inequality follows from ii, and the third inequality follows from the definition of the 2-norm.

We will find an example matrix $A \in \mathbb{R}^{2\times 2}$ for which equality holds. To do so we write down sufficient conditions under which each of the three inequalities in the above statement are equalities. Satisfying the inequalities from left to right, we impose the conditions:

1. $A\mathbf{y} = (\alpha, 0)^T$ for some α
2. $\mathbf{y} = (1, 1)^T$
3. $||A||_2 = \frac{||A\mathbf{y}||_2}{||\mathbf{y}||_2} = ||A\frac{\mathbf{y}}{||\mathbf{y}||_2}||_2$

Additionally we have have the condition

4. $||A\begin{pmatrix} 1 \\ 1 \end{pmatrix}||_\infty = ||A||_\infty = \alpha$

If we find A and $\mathbf{y}$ for which conditions 1, 2, 3, and 4 are satisfied then we will have an example for which $||A||_\infty = \sqrt{2}||A||_2$. One matrix for which condition 3 holds is $A = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \end{pmatrix}$ (since $\mathbf{y}\cdot\mathbf{u}$ is maximized for $\|\mathbf{u}\| = 1$ if $\mathbf{u}$ is parallel to $\mathbf{y}$).

Therefore, for the choice $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$, we have

$$||A||_\infty = 2,$$

and

$$||A||_2 = 2/\sqrt{2} = \sqrt{2},$$

so that

$$||A||_\infty = \sqrt{2}||A||_2.$$

iv. Let

$$\mathbf{y} = \operatorname*{argmax}_{||\mathbf{x}||_2 = 1} ||A\mathbf{x}||_2.$$

The bound holds because

$$||A\mathbf{y}||_2 \le \sqrt{n}||A\mathbf{y}||_\infty = \sqrt{n}||\mathbf{y}||_\infty\frac{||A\mathbf{y}||_\infty}{||\mathbf{y}||_\infty} \le \sqrt{n}||\mathbf{y}||_2\frac{||A\mathbf{y}||_\infty}{||\mathbf{y}||_\infty} = \sqrt{n}\frac{||A\mathbf{y}||_\infty}{||\mathbf{y}||_\infty} \le \sqrt{n}||A||_\infty$$

where the first inequality follow from ii, the second inequality follows from i, and the third inequality follows from the definition of the norm. To achieve equality, we find an example 2×2 matrix for which the inequalities are equalities. Satisfying equality for the inequalities from left to right, we impose the

conditions:
1. $A\mathbf{y} = (1,1)^T$
2. $\mathbf{y} = (1,0)^T$
3. $\frac{||A\mathbf{y}||_\infty}{||\mathbf{y}||_\infty} = ||A||_\infty$

One matrix which satisfies these conditions is

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

(b) Let $||\cdot||_a$ and $||\cdot||_b$ be two equivalent norms, so that for some constants c, d we have

$$c||\mathbf{x}||_a \le ||\mathbf{x}||_b \le d||\mathbf{x}||_a.$$

Let $\mathbf{y}$ and $\mathbf{z}$ be such that

$$||A||_a = \max_{||\mathbf{x}||_a=1} ||A\mathbf{x}||_a = ||A\mathbf{y}||_a$$
$$||A||_b = \max_{||\mathbf{x}||_b=1} ||A\mathbf{x}||_b = ||A\mathbf{z}||_b.$$

Then

$$||A||_a = ||A\mathbf{y}||_a \le \frac{1}{c}||A\mathbf{y}||_b = \frac{||\mathbf{y}||_b}{c}\frac{||A\mathbf{y}||_b}{||\mathbf{y}||_b} \le \frac{||\mathbf{y}||_b}{c}||A||_b$$
$$||A||_b = ||A\mathbf{z}||_b \le d||A\mathbf{z}||_a = d||\mathbf{z}||_a\frac{||A\mathbf{z}||_a}{||\mathbf{z}||_a} \le d||\mathbf{z}||_a||A||_a$$

Combining these, we have

$$\frac{c}{||\mathbf{y}||_b}||A||_a \le ||A||_b \le d||\mathbf{z}||_a||A||_a.$$

Conditioning and stability

4. For each of the following statements, indicate whether the statement is true or false.

$\boxed{\textbf{T}}$/**F** A problem is ill-conditioned if its solution is highly sensitive to changes in its data.

T/$\boxed{\textbf{F}}$ We can improve conditioning of a problem by switching from single to double precision arithmetic.

$\boxed{\textbf{T}}$/**F** In order to numerically solve a problem accurately, it is necessary to have both a well-conditioned problem and a stable algorithm.

$\boxed{\textbf{T}}$/**F** A condition number of 1 means the problem is well-conditioned.

5. (Heath 2.58) Suppose that the $n \times n$ matrix A is perfectly well-conditioned, i.e., cond(A) = 1. Which of the following matrices would then necessarily share this same property?

(a) cA, where c is any nonzero scalar
(b) DA, where D is a nonsingular diagonal matrix
(c) PA, where P is any permutation matrix
(d) BA, where B is any nonsingular matrix
(e) A^{-1}, the inverse of A

Solution:

We give answers, as well as explanations below. You need only give answers for full credit.

(a) cA, where c is any nonzero scalar. YES
Explanation:

$$\begin{aligned}\text{cond}(cA) &= \|cA\|\|(cA)^{-1}\| \\ &= |c|\|A\|\|c^{-1}A^{-1}\| \\ &= |c|\|A\||c^{-1}|\|A^{-1}\| \\ &= \|A\|\|A^{-1}\| = \text{cond}(A)\end{aligned}$$

(b) DA, where D is a nonsingular diagonal matrix. NO
Explanation: $\text{cond}(DA)$ can be worse than $\text{cond}(A)$. As an example, consider the case where $A = I$. Let $D = \text{diag}(d_1, d_2, \ldots, d_n)$. Then $\text{cond}(DA) = \text{cond}(D) = \max_i |d_i| / \min_i |d_i|$, which can be chosen to be large.

(c) PA, where P is any permutation matrix. YES, for the commonly used norms, but NO in general. [Ungraded question].
Explanation: $\text{cond}(A) = \|A\|\|A^{-1}\|$ and thus depends on which norm we use. For the commonly used norms, such as any induced p-norms or Frobenius norms, the norms do not change under row or column permutations. However, it is possible to define a matrix norm which does change under such permutations, which would change the condition number.

(d) BA, where B is any nonsingular matrix. NO
Explanation: For example, the case the where B is diagonal, as explained above.

(e) A^{-1}, the inverse of A. YES
Explanation: $\text{cond}(A) = \|A\|\|A^{-1}\| = \|A^{-1}\|\|(A^{-1})^{-1}\| = \text{cond}(A^{-1})$.

Orthogonality

6. (Strang I.5 1) If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal unit vectors, under what condition is $a\mathbf{u} + b\mathbf{v}$ orthogonal to $c\mathbf{u} + d\mathbf{v}$ (where a, b, c, d are scalars)? What are the lengths of those vectors (expressed using a, b, c, d)?

Solution:

We want $(a\mathbf{u} + b\mathbf{v})^T(c\mathbf{u} + d\mathbf{v}) = 0$.

$$\begin{aligned}(a\mathbf{u} + b\mathbf{v})^T(c\mathbf{u} + d\mathbf{v}) &= ac\mathbf{u}^T\mathbf{u} + ad\mathbf{u}^T\mathbf{v} + bc\mathbf{v}^T\mathbf{u} + bd\mathbf{v}^T\mathbf{v} \\ &= ac + bd\end{aligned}$$

Therefore, $a\mathbf{u} + b\mathbf{v}$ is orthogonal to $c\mathbf{u} + d\mathbf{v}$ when $ac + bd = 0$.

$$\begin{aligned}\|a\mathbf{u} + b\mathbf{v}\|_2^2 &= (a\mathbf{u} + b\mathbf{v})^T(a\mathbf{u} + b\mathbf{v}) \\ &= a^2 + b^2\end{aligned}$$

So $\|a\mathbf{u} + b\mathbf{v}\|_2 = \sqrt{a^2 + b^2}$. Similarly $\|c\mathbf{u} + d\mathbf{v}\|_2 = \sqrt{c^2 + d^2}$.

7. (Strang I.5 4) Prove this key property of every orthogonal matrix Q: $\|Q\mathbf{x}\|^2 = \|\mathbf{x}\|^2$ for every vector $\mathbf{x}$. More than this, show that $(Q\mathbf{x})^T(Q\mathbf{y}) = \mathbf{x}^T\mathbf{y}$ for every vector $\mathbf{x}$ and $\mathbf{y}$. So *lengths and angles are not changed by Q. Computations with Q never overflow!*.

Solution:

$$\begin{aligned}\|Q\mathbf{x}\|_2^2 &= (Q\mathbf{x})^T(Q\mathbf{x}) \\ &= \mathbf{x}^TQ^TQ\mathbf{x} \\ &= \mathbf{x}^T\mathbf{x} \\ &= \|\mathbf{x}\|_2^2\end{aligned}$$

$$\begin{aligned}(Q\mathbf{x})^T(Q\mathbf{y}) &= \mathbf{x}^TQ^TQ\mathbf{y} \\ &= \mathbf{x}^T\mathbf{y}\end{aligned}$$

8. (Strang I.5 6) A permutation matrix has the same columns as the identity matrix (in some order). Explain why this permutation matrix and every permutation is orthogonal:

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Solution:

$$\begin{aligned} PP^T &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= I \end{aligned}$$

Generally, let P be a permutation matrix with $P_{ij} = 1$. And let $y = Px$. Then $y_i = (Px)_i = x_j$, i.e. P takes the jth element of x and permutes it into the ith position of y. Since $P_{ij} = 1$, we know $(P^T)_{ji} = 1$ as well. This means that $(P^T y)_j = y_i$. Therefore $((P^T P)x)_j = (P^T(Px))_j = (Px)_i = x_j$. Therefore $P^T P = I$.

Alternatively, P_{ij} permutes element j into position i. P^T_{ji} permutes element at i into j. Therefore P^T undoes action of P, so $P^T P = I$.

9. Let $A \in \mathbb{R}^{n\times m}$. Show that range$(A)$ is orthogonal to nullspace(A^T) . I.e., show that for any $\mathbf{y} \in$ range$(A), \mathbf{z} \in$ nullspace(A), $\mathbf{y}^T\mathbf{z} = 0$.

Solution:

range$(A) = \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{y} = A\mathbf{x}, \mathbf{x} \in \mathbb{R}^m\}$

nullspace$(A^T) = \{\mathbf{z} \in \mathbb{R}^m | A^T\mathbf{z} = \mathbf{0}\}$

Given any $\mathbf{y} \in$ range(A), with $\mathbf{y} = A\mathbf{x}$, and any $\mathbf{z} \in$ nullsapce(A^T), we have

$$\mathbf{y}^T\mathbf{z} = \mathbf{x}^T A^T\mathbf{z} = \mathbf{x}^T(A^T\mathbf{z}) = \mathbf{x}^T\mathbf{0} = 0.$$

Therefore range(A) is orthogonal to nullspace(A^T).

Projections

10. Given a vector $\mathbf{v} \in \mathbb{R}^n$ and an orthonomal set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ in $\mathbb{R}^n$, find $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ that minimize $\|\mathbf{v} - \sum_{i=1}^k \alpha_i\mathbf{u}_i\|_2^2$.

Solution:

Note that since the $\mathbf{u}_i$'s are orthonomal, $\mathbf{u}_i^T\mathbf{u}_j = \delta_{ij}$.

We first simplify the objective we wish to minimize:

$$\begin{aligned} E(\alpha_1, ..., \alpha_k) &= \|\mathbf{v} - \sum_{i=1}^k \alpha_i\mathbf{u}_i\|_2^2 \\ &= (\mathbf{v} - \sum_{i=1}^k \alpha_i\mathbf{u}_i)^T(\mathbf{v} - \sum_{i=1}^k \alpha_i\mathbf{u}_i) \\ &= \mathbf{v}^T\mathbf{v} - 2\sum_{i=1}^k \alpha_i\mathbf{u}_i^T\mathbf{v} + (\sum_{i=1}^k \alpha_i\mathbf{u}_i^T)(\sum_{i=1}^k \alpha_i\mathbf{u}_i) \\ &= \|\mathbf{v}\|_2^2 - 2\sum_{i=1}^k \alpha_i\mathbf{u}_i^T\mathbf{v} + \sum_{i=1}^k \alpha_i^2 \end{aligned}$$

Differentiating w.r.t. α_j, we get

$$\frac{\partial E(\alpha_1, ..., \alpha_k)}{\partial \alpha_j} = -2\mathbf{u}_j^T\mathbf{v} + 2\alpha_j$$

The minimum occurs when this quantity is zero:

$$\begin{aligned} 0 = \frac{\partial E}{\partial \alpha_j} &= -2\mathbf{u}_j^T\mathbf{v} + 2\alpha_j \\ \alpha_j &= \mathbf{u}_j^T\mathbf{v} \end{aligned}$$

So the α_i are simply the projected lengths of $\mathbf{v}$ onto the $\mathbf{u}_i$.

Therefore, the closest vector under the 2-norm to $\mathbf{v}$ in the subspace spanned by $\{\mathbf{u}_i, ..., \mathbf{u}_k\}$ is $\mathbf{u}_1\mathbf{u}_1^T\mathbf{v} + \mathbf{u}_2\mathbf{u}_2^T\mathbf{v} + ... + \mathbf{u}_k\mathbf{u}_k^T\mathbf{v}$

11. Let P be a projection matrix (i.e., $P^2 = P$).

(a) Prove that $I - P$ is also a projection matrix. This is called the *complementary projector* to P.

Solution:

To prove that $I - P$ is a projection matrix, we just need to show that it is idempotent:

$$\begin{aligned} (I - P)(I - P) &= I - 2P + P^2 \\ &= I - 2P + P \\ &= I - P \end{aligned}$$

(b) Prove that if P is a symmetric matrix, then $P\mathbf{x}$ is orthogonal to $(I - P)\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Such P is called an *orthogonal projection*. (Note also that the converse holds, i.e., $P\mathbf{x} \perp (I - P)\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n \Rightarrow P = P^T$.)

Solution:

Let $P = P^T$, then

$$\begin{aligned} (P\mathbf{x})^T(I - P)\mathbf{y} &= \mathbf{x}^T P^T (I - P)\mathbf{y} \\ &= \mathbf{x}^T P^T \mathbf{y} - \mathbf{x}^T P^T P\mathbf{y} \\ &= \mathbf{x}^T P\mathbf{y} - \mathbf{x}^T P^2\mathbf{y} \\ &= \mathbf{x}^T P\mathbf{y} - \mathbf{x}^T P\mathbf{y} \\ &= 0 \end{aligned}$$

Proof of the converse (not graded):

$P\mathbf{x} \perp (I - P)\mathbf{y}$, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$\begin{aligned} 0 &= \mathbf{x}^T P^T (I - P)\mathbf{y} \\ \mathbf{x}^T P^T \mathbf{y} &= \mathbf{x}^T P^T P\mathbf{y} \end{aligned}$$

Since $\mathbf{x}^T P^T \mathbf{y} = \mathbf{x}^T P^T P\mathbf{y}$ holds for all $\mathbf{y}, \mathbf{x} \in \mathbb{R}^n$, it must be that

$$P^T = P^T P.$$

(Why? We can choose $\mathbf{x} = \mathbf{e}_i$ and $\mathbf{y} = \mathbf{e}_j$, so that $(\mathbf{x}^T P^T \mathbf{y}) = P_{ij}^T$, $(\mathbf{x}^T P^T P\mathbf{y}) = (P^T P)_{ij}$ and therefore $P_{ij}^T = (P^T P)_{ij}\ \ \forall i, j$.) Transposing both side, we also have

$$P = P^T P.$$

Therefore $P = P^T$.