2. Constrained Least Squares Optimization

(a) Direct Proof

Consider the corresponding eigenvector, $\vec{v_1}$, of eigenvalue λ_1. By definition, we have that $\mathbf{A}^T\mathbf{A}\vec{v_1} = \lambda_1\vec{v_1}$, which gives that

Then, by definition of the 2-norm, we have that

$$\|\mathbf{A}\vec{v_1}\|^2 = \vec{v_1}^T\mathbf{A}^T\mathbf{A}\vec{v_1} = \vec{v_1}^T\lambda_1\vec{v_1} = \lambda_1\vec{v_1}^T\vec{v_1} = \lambda_1 \cdot \|\vec{v_1}\|^2$$

Since the matrix $\mathbf{A}$ is full rank, so it has a trivial nullspace, i.e. for non-trivial vector $\vec{v_1}$, we have that $\mathbf{A}\vec{v_1} \neq \vec{0}$ and $\|\vec{v_1}\| \neq$, and thus, $\|\mathbf{A}\vec{v_1}\| \neq 0$. By the nature of squares, so

$$\|\mathbf{A}\vec{v_1}\|^2 > 0, \text{and } \|\vec{v_1}\|^2 > 0$$

Therefore, we have that

$$\lambda_1 = \frac{\|\mathbf{A}\vec{v_1}\|^2}{\|\vec{v_1}\|^2} > 0$$

as desired, i.e. all the eigenvalues are strictly positive.

Q.E.D.

(b) Direct Proof

Given the two equations:

$$\mathbf{A}^T\mathbf{A}\vec{v_k} = \lambda_k\vec{v_k} \tag{1}$$

$$\vec{v_l}^T\mathbf{A}^T\mathbf{A} = \lambda_l\vec{v_l}^T \tag{2}$$

Premultiply Equation 1 with $\vec{v_l}^T$, and postmultiply Equation 2 with $\vec{v_k}$ gives us the two following equations:

$$\vec{v_l}^T\mathbf{A}^T\mathbf{A}\vec{v_k} = \vec{v_l}^T\lambda_k\vec{v_k}$$
$$\vec{v_l}^T\mathbf{A}^T\mathbf{A}\vec{v_k} = \lambda_l\vec{v_l}^T\vec{v_k}$$

Thus, we can conclude that:

$$\vec{v_l}^T\lambda_k\vec{v_k} = \vec{v_l}^T\mathbf{A}^T\mathbf{A}\vec{v_k} = \lambda_l\vec{v_l}^T\vec{v_k}$$

which, after rearranging the terms to put constant (eigenvalue) as first, would give us:

$$\lambda_k\vec{v_l}^T\vec{v_k} = \lambda_l\vec{v_l}^T\vec{v_k} \tag{3}$$

Since we have that $\lambda_k \neq \lambda_l$, and that we proved in part (a) that all eigenvalues are strictly positive, which gives us that $\lambda_k, \lambda_l \neq 0$, so, for Equation 3 to hold, it's necessary that:

$$\vec{v_l}^T\vec{v_k} = 0$$

$$\implies \langle\vec{v_l}, \vec{v_k}\rangle = 0$$

which gives us that $\vec{v_k}$ and $\vec{v_l}$ are orthogonal, as desired. Q.E.D.

(c) (i) $\alpha_n = \vec{v_n}^T \vec{x}$; (ii) Direct Proof

(i) To determine the n^th coefficient α_n, consider as we multiply $\vec{v_n}^T$ by $\vec{x}$. Given that $\vec{x} = \sum\limits_{n=1}^{N} \alpha_n \vec{v_n}$, so we have that:

$$\vec{v_n}^T \vec{x} = \vec{v_n}^T \cdot \Big(\sum_{k=1}^{N} \alpha_k \vec{v_k} \Big) =$$

$$= \vec{v_n}^T \alpha_1 \vec{v_1} + \vec{v_n}^T \alpha_2 \vec{v_2} + \cdots + \vec{v_n}^T \alpha_n \vec{v_n} + \cdots + \vec{v_n}^T \alpha_N \vec{v_N} = \alpha_1 \vec{v_n}^T \vec{v_1} + \alpha_2 \vec{v_n}^T \vec{v_2} + \cdots + \alpha_n \vec{v_n}^T \vec{v_n} + \cdots + \alpha_N \vec{v_n}^T \vec{v_N}$$

Since $\vec{v_1}, \cdots, \vec{v_N}$ form an orthonormal basis and they're all unit length, so by definition, for any $i, j, n \in [1, N], i \neq j$, we have that $\langle \vec{v_i}, \vec{v_j} \rangle = 0$, and that $\langle \vec{v_n}, \vec{v_n} \rangle = \vec{v_n}^T \vec{v_n} = \|\vec{v_n}\|^2 = 1$. Thus, we have that:

$$\Longrightarrow \vec{v_n}^T \vec{x} = 0 + 0 + \cdots \alpha_n \cdot 1 + 0 + \cdots + 0 = \alpha_n$$

Therefore, the n^th coefficient is:

$$\alpha_n = \vec{v_n}^T \vec{x}$$

(ii) Using the fact that $\vec{x} = \sum\limits_{n=1}^{N} \alpha_n \vec{v_n}$, consider

$$\|\vec{x}\|^2 = \Big(\sum_{n=1}^{N} \alpha_n \vec{v_n} \Big)^2$$

Expand the square of the summation and we get:

$$\|\vec{x}\|^2 = (\alpha_1 \vec{v_1} \cdot \alpha_1 \vec{v_1} + \cdots + \alpha_1 \vec{v_1} \cdot \alpha_N \vec{v_N}) + \cdots\cdots + (\alpha_N \vec{v_N} \cdot \alpha_1 \vec{v_1} + \cdots + \alpha_N \vec{v_N} \cdot \alpha_N \vec{v_N})$$

$$\Longrightarrow \|\vec{x}\|^2 = (\alpha_1^2 \vec{v_1} \vec{v_1} + \alpha_1 \alpha_2 \vec{v_1} \vec{v_2} + \cdots + \alpha_1 \alpha_N \vec{v_1} \vec{v_N}) + \cdots\cdots + (\alpha_N \alpha_1 \vec{v_N} \vec{v_1} + \cdots + \alpha_N^2 \vec{v_N} \vec{v_N})$$

As similarly demonstrated in part (i), for any $i, j, n \in [1, N], i \neq j$, we have that $\langle \vec{v_i}, \vec{v_j} \rangle = 0$, and that $\langle \vec{v_n}, \vec{v_n} \rangle = \vec{v_n}^T \vec{v_n} = \|\vec{v_n}\|^2 = 1$. Thus, we can simplify the result above to:

$$\|\vec{x}\|^2 = (\alpha_1^2 \cdot 1 + 0 + \cdots + 0) + (0 + \alpha_2^2 \cdot 1 + \cdots + 0) + (0 + \cdots + \alpha_N^2 \cdot 1) = \alpha_1^2 + \cdots + \alpha_N^2$$

Since $\vec{x}$ is a unit vector, so the left side of the equation can be simplified to $\|\vec{x}\|^2 = 1$, and the right side could be rewritten as $\sum\limits_{n=1}^{N} \alpha_n^2$.

Therefore, we can conclude that:

$$\sum_{n=1}^{N} \alpha_n^2 = 1$$

as desired.

Q.E.D.

(d) $\|\mathbf{A}\vec{x}\|^2 = \sum_{n=1}^{N} \alpha_n^2 \lambda_n;\ \vec{\tilde{x}} = \vec{v_1};\ \|\mathbf{A}\vec{x}\| = \lambda_1$

Using the given equations, setup and results from previous parts, we have that:

$$\|\mathbf{A}\vec{x}\|^2 = \vec{x}^T \mathbf{A}^T \mathbf{A} \vec{x}$$

$$\mathbf{A}^T \mathbf{A}\, \vec{v_k} = \lambda_k \vec{v_k} \qquad \text{for all } k \in \{1, \dots, N\}$$

$$\vec{x} = \sum_{n=1}^{N} \alpha_n \vec{v_n}, \text{ and so } \vec{x}^T = \sum_{n=1}^{N} \alpha_n \vec{v_n}^T$$

Also, as similarly demonstrated in part (i), for any $i, j, n \in [1, N], i \neq j$, we have that $\langle \vec{v_i}, \vec{v_j} \rangle = 0$, and that $\langle \vec{v_n}, \vec{v_n} \rangle = \vec{v_n}^T \vec{v_n} = \|\vec{v_n}\|^2 = 1$.

Thus, we can rewrite $\|\mathbf{A}\vec{x}\|^2$ as:

$$\|\mathbf{A}\vec{x}\|^2 = \Big(\sum_{n=1}^{N} \alpha_n \vec{v_n}^T\Big) \cdot \mathbf{A}^T \mathbf{A} \cdot \Big(\sum_{n=1}^{N} \alpha_n \vec{v_n}\Big) \tag{4}$$

Now, since $\mathbf{A}^T \mathbf{A}\, \vec{v_k} = \lambda_k \vec{v_k}$ for all $k \in \{1, \dots, N\}$, so we have

$$\mathbf{A}^T \mathbf{A}\, \alpha_k \vec{v_k} = \alpha_k \lambda_k \vec{v_k}, \text{ for all } k \in \{1, \dots, N\}$$

Thus,

$$\mathbf{A}^T \mathbf{A} \cdot \Big(\sum_{n=1}^{N} \alpha_n \vec{v_n}\Big) = \sum_{n=1}^{N} \alpha_n \lambda_n \vec{v_n}$$

Thus, we can further simply Eq. (4) as:

$$\|\mathbf{A}\vec{x}\|^2 = \Big(\sum_{n=1}^{N} \alpha_n \vec{v_n}^T\Big) \cdot \sum_{n=1}^{N} \alpha_n \lambda_n \vec{v_n}$$

Expanding the product of the two summations into individual terms gives us that $\|\mathbf{A}\vec{x}\|^2 = (\alpha_1 \vec{v_1}^T \cdot \alpha_1 \lambda_1 \vec{v_1} + \alpha_1 \vec{v_1}^T \cdot \alpha_2 \lambda_2 \vec{v_2} + \dots + \alpha_1 \vec{v_1}^T \cdot \alpha_N \lambda_N \vec{v_N}) + (\alpha_2 \vec{v_2}^T \cdot \alpha_1 \lambda_1 \vec{v_1} + \alpha_2 \vec{v_2}^T \cdot \alpha_2 \lambda_2 \vec{v_2} + \dots + \alpha_2 \vec{v_2}^T \cdot \alpha_N \lambda_N \vec{v_N}) + \dots\dots + (\alpha_N \vec{v_N}^T \cdot \alpha_1 \lambda_1 \vec{v_1} + \alpha_N \vec{v_N}^T \cdot \alpha_2 \lambda_2 \vec{v_2} + \dots + \alpha_N \vec{v_N}^T \cdot \alpha_N \lambda_N \vec{v_N})$, which, by pulling constants to the front, we could rewrite as:

$$\|\mathbf{A}\vec{x}\|^2 = (\alpha_1 \alpha_1 \lambda_1 \cdot \vec{v_1}^T \vec{v_1} + \dots + \alpha_1 \alpha_N \lambda_N \cdot \vec{v_1}^T \vec{v_N}) \ + \ \dots \ + \ (\alpha_N \alpha_1 \lambda_1 \cdot \vec{v_N}^T \vec{v_1} + \dots + \alpha_N \alpha_N \lambda_N \cdot \vec{v_N}^T \vec{v_N})$$

Using the orthonormal property (details included above), so we can further simply our result to be:

$$\|\mathbf{A}\vec{x}\|^2 = \alpha_1 \alpha_1 \lambda_1 \cdot 1 + \alpha_2 \alpha_2 \lambda_2 \cdot 1 + \dots + \alpha_N \alpha_N \lambda_N \cdot 1 = \sum_{n=1}^{N} \alpha_n^2 \lambda_n$$

To minimize $\|\mathbf{A}\vec{x}\|^2 = \alpha_1^2 \lambda_1 + \dots + \alpha_N^2 \lambda_N$ with the constraint that $\sum_{n=1}^{N} \alpha_n^2 = 1$, and also since we set the λ's such that $\lambda_1 < \dots < \lambda_N$, so the minimum occurs when $\alpha_1^2 = 1,\ \alpha_2^2 = \dots = \alpha_N^2 = 0$. Thus, in this case, the $\vec{\tilde{x}}$ we're looking for is:

$$\vec{\tilde{x}} = 1 \cdot \vec{v_1} + 0 \cdot \vec{v_2} + \dots + 0 \cdot \vec{v_N} = \vec{v_1}$$

Therefore, with $\vec{v_1}$ being a unit vector, so we have:

$$\|\mathbf{A}\vec{x}\| = \|\lambda_1 \vec{v_1}\| = \lambda_1 \cdot \|\vec{v_1}\| = \lambda_1$$