I. Linear Algebra

I.6. The Spectral Decomposition

Lecture based on

https://github.com/gwthomas/math4ml (Garrett Thomas, 2018)

https://mml-book.github.io/ (Deisenroth et al., 2020, Mathematics for Machine Learning)

Prof. Dr. Christoph Lippert

Digital Health & Machine Learning

A matrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is said to be **orthogonal** if its columns are pairwise orthonormal.

This definition implies that

$$\mathbf{Q}^\top \mathbf{Q} = \mathbf{Q}\mathbf{Q}^\top = \mathbf{I}$$

or equivalently, $\mathbf{Q}^\top = \mathbf{Q}^{-1}$.

A nice thing about orthogonal matrices is that they preserve inner products:

$$(\mathbf{Q}\mathbf{x})^\top (\mathbf{Q}\mathbf{y}) = \mathbf{x}^\top \mathbf{Q}^\top \mathbf{Q}\mathbf{y} = \mathbf{x}^\top \mathbf{I}\mathbf{y} = \mathbf{x}^\top \mathbf{y}$$

A direct result of this fact is that they also preserve 2-norms:

$$\|\mathbf{Q}\mathbf{x}\|_2 = \sqrt{(\mathbf{Q}\mathbf{x})^\top (\mathbf{Q}\mathbf{x})} = \sqrt{\mathbf{x}^\top \mathbf{x}} = \|\mathbf{x}\|_2$$

Therefore multiplication by an orthogonal matrix can be considered as a transformation that preserves length, but may rotate or reflect the vector about the origin.

The **trace** of a square matrix is the sum of its diagonal entries:

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{n} A_{ii}$$

The trace has several nice algebraic properties:

(i) $\operatorname{tr}(\mathbf{A} + \mathbf{B}) = \operatorname{tr}(\mathbf{A}) + \operatorname{tr}(\mathbf{B})$

(ii) $\operatorname{tr}(\alpha\mathbf{A}) = \alpha \operatorname{tr}(\mathbf{A})$

(iii) $\operatorname{tr}\left(\mathbf{A}^\top\right) = \operatorname{tr}(\mathbf{A})$

(iv) $\operatorname{tr}(\mathbf{ABCD}) = \operatorname{tr}(\mathbf{BCDA}) = \operatorname{tr}(\mathbf{CDAB}) = \operatorname{tr}(\mathbf{DABC})$

The first three properties follow readily from the definition.

The last is known as **invariance under cyclic permutations**.

Note that the matrices cannot be reordered arbitrarily, for example $\mathrm{tr}(\mathbf{ABCD}) \neq \mathrm{tr}(\mathbf{BACD})$ in general.

Also, there is nothing special about the product of four matrices – analogous rules hold for more or fewer matrices.

Interestingly, the trace of a matrix is equal to the sum of its eigenvalues (repeated according to multiplicity):

$$\mathrm{tr}(\mathbf{A}) = \sum_i \lambda_i(\mathbf{A})$$

A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be **symmetric** if it is equal to its own transpose ($\mathbf{A} = \mathbf{A}^\top$), meaning that $A_{ij} = A_{ji}$ for all (i, j).

This definition seems harmless enough but turns out to have some strong implications.

We summarize the most important of these as

Theorem (Spectral Theorem)

If $\mathbf{A} \in \mathbb{R}^{N \times N}$ is symmetric, then there exists an orthonormal basis for $\mathbb{R}^N$ consisting of eigenvectors of $\mathbf{A}$.

The practical application of this theorem is a particular factorization of symmetric matrices, referred to as the **eigendecomposition** or **spectral decomposition**.

Denote the orthonormal basis of eigenvectors $\mathbf{q}_1, \ldots, \mathbf{q}_n$ and their eigenvalues $\lambda_1, \ldots, \lambda_n$.

Let $\mathbf{Q}$ be an orthogonal matrix with $\mathbf{q}_1, \ldots, \mathbf{q}_n$ as its columns, and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

Since by definition $\mathbf{A}\mathbf{q}_i = \lambda_i \mathbf{q}_i$ for every i, the following relationship holds:

$$\mathbf{AQ} = \mathbf{Q\Lambda}$$

Right-multiplying by $\mathbf{Q}^\top$, we arrive at the decomposition

$$\mathbf{A} = \mathbf{Q\Lambda Q}^\top$$

Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ be a symmetric matrix.

The expression $\mathbf{x}^\top \mathbf{A} \mathbf{x}$ is called a **quadratic form**.

The **Rayleigh quotient**

$$R_{\mathbf{A}}(\mathbf{x}) = \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}}$$

The Rayleigh quotient has a couple of important properties which the reader can (and should!) easily verify from the definition:

i **Scale invariance**: for any vector $\mathbf{x} \neq \mathbf{0}$ and any scalar $\alpha \neq 0$, $R_{\mathbf{A}}(\mathbf{x}) = R_{\mathbf{A}}(\alpha\mathbf{x})$.

ii If $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ, then $R_{\mathbf{A}}(\mathbf{x}) = \lambda$.

The Rayleigh quotient is bounded by the largest and smallest eigenvalues of $\mathbf{A}$.

Proposition

For any $\mathbf{x}$ *such that* $\|\mathbf{x}\|_2 = 1$,

$$\lambda_{\min}(\mathbf{A}) \leq \mathbf{x}^\top \mathbf{A}\mathbf{x} \leq \lambda_{\max}(\mathbf{A})$$

with equality if and only if $\mathbf{x}$ *is a corresponding eigenvector.*

Proof.

We show only the $\max$ case because the argument for the $\min$ case is entirely analogous.
Since $\mathbf{A}$ is symmetric, we can decompose it as $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$. Then use the change of variable $\mathbf{y} = \mathbf{Q}^\top\mathbf{x}$, noting that the relationship between $\mathbf{x}$ and $\mathbf{y}$ is one-to-one and that $\|\mathbf{y}\|_2 = 1$ since $\mathbf{Q}$ is orthogonal.
Hence

$$\max_{\|\mathbf{x}\|_2=1} \mathbf{x}^\top\mathbf{A}\mathbf{x} = \max_{\|\mathbf{y}\|_2=1} \mathbf{y}^\top\mathbf{\Lambda}\mathbf{y} = \max_{y_1^2+\cdots+y_n^2=1} \sum_{i=1}^{n} \lambda_i y_i^2$$

Written this way, it is clear that $\mathbf{y}$ maximizes this expression exactly if and only if it satisfies $\sum_{i\in I} y_i^2 = 1$ where $I = \{i : \lambda_i = \max_{j=1,\ldots,n} \lambda_j = \lambda_{\max}(\mathbf{A})\}$ and $y_j = 0$ for $j \notin I$.
That is, I contains the index or indices of the largest eigenvalue.
In this case, the maximal value of the expression is

$$\sum_{i=1}^{n} \lambda_i y_i^2 = \sum_{i\in I} \lambda_i y_i^2 = \lambda_{\max}(\mathbf{A}) \sum_{i\in I} y_i^2 = \lambda_{\max}(\mathbf{A})$$

To be contiued . . . $\square$

... continued.

Then writing $\mathbf{q}_1, \ldots, \mathbf{q}_n$ for the columns of $\mathbf{Q}$, we have

$$\mathbf{x} = \mathbf{Q}\mathbf{Q}^\top \mathbf{x} = \mathbf{Q}\mathbf{y} = \sum_{i=1}^{n} y_i \mathbf{q}_i = \sum_{i \in I} y_i \mathbf{q}_i$$

where we have used the matrix-vector product identity.
Recall that $\mathbf{q}_1, \ldots, \mathbf{q}_n$ are eigenvectors of $\mathbf{A}$ and form an orthonormal basis for $\mathbb{R}^n$.
Therefore by construction, the set $\{\mathbf{q}_i : i \in I\}$ forms an orthonormal basis for the eigenspace of $\lambda_{\max}(\mathbf{A})$.
Hence $\mathbf{x}$, which is a linear combination of these, lies in that eigenspace and thus is an eigenvector of $\mathbf{A}$ corresponding to $\lambda_{\max}(\mathbf{A})$.
We have shown that $\max_{\|\mathbf{x}\|_2=1} \mathbf{x}^\top \mathbf{A}\mathbf{x} = \lambda_{\max}(\mathbf{A})$, from which we have the general inequality $\mathbf{x}^\top \mathbf{A}\mathbf{x} \le \lambda_{\max}(\mathbf{A})$ for all unit-length $\mathbf{x}$. $\square$

By the scale invariance of the Rayleigh quotient, we immediately have as a corollary (since $\mathbf{x}^\top \mathbf{A}\mathbf{x} = R_{\mathbf{A}}(\mathbf{x})$ for unit $\mathbf{x}$)

Theorem (Min-max theorem)

For all $\mathbf{x} \neq \mathbf{0}$,

$$\lambda_{\min}(\mathbf{A}) \leq R_{\mathbf{A}}(\mathbf{x}) \leq \lambda_{\max}(\mathbf{A})$$

with equality if and only if $\mathbf{x}$ is a corresponding eigenvector.