0.1 Proof of theorem 0.1

Our relaxed objective becomes:

$$\begin{aligned} \underset{M}{\text{minimize}} \quad & f_S' := \mathrm{Tr}(\mathbf{M}^T\mathbf{L}\mathbf{M}), \\ \text{subject to} \quad & \mathbf{M}^T\mathbf{D}\mathbf{M} = \mathbf{1}^{k\times k} \text{ and } \mathbf{M} \in \mathbb{R}^{n\times k}. \end{aligned}$$

The optimal solution to f_S' is more tractable to compute.

Theorem 0.1. *The optimal solution to f_S' is the matrix $\mathbf{U}$ with the first k eigenvectors[1] of the generalized eigenvalue problem $\mathbf{Lu} = \lambda\mathbf{Du}$ as its columns.*

Originally stated without proof in [2], we construct a full proof of the theorem here. It requires finding a lower bound for $\mathrm{Tr}(\mathbf{M}^T\mathbf{L}\mathbf{M})$, and showing that $\mathbf{M} := \mathbf{U}$ achieves this bound. For a brief mathematical background, we again refer to appendix A.

Lemma 0.2 (Courant-Fischer Min-Max Theorem)**.** *Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be some symmetric matrix with eigenvalues*

$$\lambda_1 \leq \cdots \leq \lambda_n.$$

Then,

$$\lambda_d = \min_{\substack{S\subseteq\mathbb{R}^n \\ \dim(S)=d}} \max_{\substack{\mathbf{x}\in S \\ \mathbf{x}\neq\mathbf{0_n}}} R_{\mathbf{A}}(\mathbf{x}),$$

where $R_{\mathbf{A}}(\mathbf{x})$ denotes the Rayleigh-Ritz quotient [1], defined as

$$\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}},$$

for any $1 \leq d \leq n$.

Proof. Let $\mathbf{u}_1, \dots, \mathbf{u}_n \in \mathbb{R}^n$ be an orthonormal set of eigenvectors of $\mathbf{A}$ corresponding to eigenvalues $\lambda_1, \dots, \lambda_n$, implying $\|\mathbf{u}_d\|^2 = 1$ for $1 \leq d \leq n$. Since $\mathbf{A}$ is symmetric, this set exists (see appendix A.1.2).

Now let $S_d := \mathrm{span}(\mathbf{u}_d, \dots, \mathbf{u}_n)$ and let $S \subseteq \mathbb{R}^n$ be an arbitrary set with $\dim(S) = d$. If $S_d \cap S = \{\mathbf{0}^n\}$, then by lemma A.4, $S_d + S$ would be a subspace of $\mathbb{R}^n$ of dimension $d + (n - d + 1) = n + 1$, which is impossible by lemma A.5. Therefore, $S_d \cap S \neq \{\mathbf{0}^n\}$. Now let $\mathbf{s} \in S_d \cap S$ such that $\mathbf{s} \neq \mathbf{0}^n$. Since $\mathbf{s} \in S_d$, we can write $\mathbf{s}$ as a linear combination $\mathbf{s} = \sum_{j=d}^n \alpha_j \mathbf{u}_j$, for scalars $\alpha_j \in \mathbb{R}$. Then, we see

$$R_{\mathbf{A}}(\mathbf{s}) = \frac{\mathbf{s}^T\mathbf{A}\mathbf{s}}{\mathbf{s}^T\mathbf{s}} = \frac{\sum_{j=d}^n \alpha_j^2\lambda_j}{\sum_{j=d}^n \alpha_j^2} \geq \frac{\sum_{j=d}^n \alpha_j^2\lambda_d}{\sum_{j=d}^n \alpha_j^2} = \lambda_d,$$

where the inequality follows from $\lambda_d \leq \cdots \leq \lambda_n$. Therefore,

$$\max_{\substack{\mathbf{x}\in S \\ x\neq\mathbf{0}^n}} R_{\mathbf{A}}(\mathbf{x}) \geq \lambda_d.$$

Since S was chosen as an arbitrary subspace of $\mathbb{R}^n$ with dimension d, this statement holds for all such sets.

In particular, it holds for $S_1 := \mathrm{span}(\mathbf{u}_1, \dots, \mathbf{u}_d)$. Again, using lemma A.4, we can pick $\mathbf{0}^n \neq \mathbf{s} \in S_1 \cap S_d$, and write $\mathbf{s} = \sum_{j=1}^d \beta_j \mathbf{u}_j$, for scalars $\beta_j \in \mathbb{R}$. Analogously, we get

$$R_{\mathbf{A}}(\mathbf{s}) = \frac{\mathbf{s}^T\mathbf{A}\mathbf{s}}{\mathbf{s}^T\mathbf{s}} = \frac{\sum_{j=1}^n \beta_j^2\lambda_j}{\sum_{j=1}^n \beta_j^2} \leq \frac{\sum_{j=1}^n \beta_j^2\lambda_d}{\sum_{j=1}^n \beta_j^2} = \lambda_d,$$

where the inequality follows from $\lambda_1 \leq \cdots \leq \lambda_d$. Combining the two inequalities, we get

$$\min_{\substack{S\subseteq\mathbb{R}^n \\ \dim(S)=d}} \max_{\substack{\mathbf{x}\in S \\ \mathbf{x}\neq\mathbf{0}^n}} R_{\mathbf{A}}(\mathbf{x}) = \lambda_d,$$

which is what we wanted to show. □

[1] By 'the first k eigenvectors' we mean eigenvectors corresponding to the k smallest eigenvalues.

Corollary 0.3. *We have*

$$\lambda_n = \max_{\substack{\mathbf{x}\in\mathbb{R}^n \\ \mathbf{x}\neq\mathbf{0}^n}} R_{\mathbf{A}}(\mathbf{x})$$

Proof. Any subspace S of $\mathbb{R}^n$ of dimension n is equal to $\mathbb{R}^n$ itself. The statement now follows from lemma 0.2. □

Corollary 0.4. *Let* $\mathbf{A}$ *as in lemma 0.2. Then*

$$\lambda_d = \max_{\substack{S\subseteq\mathbb{R}^n \\ \dim(S)=n-d+1}} \min_{\substack{\mathbf{x}\in S \\ \mathbf{x}\neq\mathbf{0}^n}} R_{\mathbf{A}}(\mathbf{x})$$

Proof. Analogous as in lemma 0.2, by swapping S_d and S_1. □

Lemma 0.5 (Part of Cauchy's Interlacing Theorem)**.** *Suppose that* $1 \leq k \leq n$*. Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *be symmetric, partitioned as*

$$\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{D} \end{bmatrix}, \text{ with } \mathbf{B}\in\mathbb{R}^{k\times k}, \mathbf{D}\in\mathbb{R}^{n-k\times n-k}, \mathbf{C}\in\mathbb{R}^{k\times n-k},$$

and let the $\lambda_1^{\mathbf{A}} \leq \cdots \leq \lambda_n^{\mathbf{A}}$ *and* $\lambda_1^{\mathbf{B}} \leq \cdots \leq \lambda_k^{\mathbf{B}}$ *be the eigenvalues of* $\mathbf{A}$ *and* $\mathbf{B}$*. Then,* $\lambda_d^{\mathbf{A}} \leq \lambda_d^{\mathbf{B}}$ *for* $1 \leq d \leq k$*.*

Proof. From corollary 0.4, we have

$$\lambda_d^{\mathbf{A}} = \max_{\substack{S\subseteq\mathbb{R}^n \\ \dim(S)=n-d+1}} \min_{\substack{\mathbf{x}\in S \\ \mathbf{x}\neq\mathbf{0}^n}} R_{\mathbf{A}}(\mathbf{x}).$$

Similarly, we get

$$\lambda_d^{\mathbf{B}} = \max_{\substack{S\subseteq\mathbb{R}^k \\ \dim(S)=(n-k)-d+1}} \min_{\substack{\mathbf{y}\in S \\ \mathbf{y}\neq\mathbf{0}^n}} R_{\mathbf{B}}(\mathbf{y}).$$

Now for any $\mathbf{y} \in \mathbb{R}^k$ define $\bar{\mathbf{y}} \in \mathbb{R}^n$ which has the same elements as $\mathbf{y}$ for the first k dimensions, and zero for the last $n-k$ dimensions. Then, $\mathbf{y}^T\mathbf{B}\mathbf{y} = \bar{\mathbf{y}}^T\mathbf{A}\bar{\mathbf{y}}$ and also $\mathbf{y}^T\mathbf{y} = \bar{\mathbf{y}}^T\bar{\mathbf{y}}$, whence $R_{\mathbf{B}}(\mathbf{y}) = R_{\mathbf{A}}(\bar{\mathbf{y}})$. We can now rewrite $\lambda_d^{\mathbf{B}}$ as

$$\lambda_d^{\mathbf{B}} = \max_{\substack{S\subseteq\mathbb{R}^n \\ \dim(S)=n-d+1}} \min_{\substack{\bar{\mathbf{y}}\in S,\dagger \\ \mathbf{y}\neq\mathbf{0}^n}} R_{\mathbf{A}}(\bar{\mathbf{y}}),$$

where † refers to the condition on $\bar{\mathbf{y}}$ that its last $n-k$ elements have to equal zero. Clearly, this is the same statement as $\lambda_d^{\mathbf{A}}$, but with the extra condition †. If we substitute $\mathbf{x} := \bar{\mathbf{y}}$, we thus get

$$\lambda_d^{\mathbf{A}} = \max_{\substack{S\subseteq\mathbb{R}^n \\ \dim(S)=n-d+1}} \min_{\substack{\mathbf{x}\in S \\ \mathbf{x}\neq\mathbf{0}^n}} R_{\mathbf{A}}(\mathbf{x}) \leq \max_{\substack{S\subseteq\mathbb{R}^n \\ \dim(S)=n-d+1}} \min_{\substack{\mathbf{x}\in S,\dagger \\ \mathbf{x}\neq\mathbf{0}^n}} R_{\mathbf{A}}(\mathbf{x}) = \lambda_d^{\mathbf{B}}.$$

□

Corollary 0.6 (part of the Poincaré Separation Theorem)**.** *Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *be symmetric and suppose that* $1 \leq k \leq n$*. Let* $\mathbf{u}_1, \ldots, \mathbf{u}_k \in \mathbb{R}^n$ *be an orthonormal set of vectors and let* $\mathbf{V} \in \mathbb{R}^{n\times k}$ *be the (orthogonal) matrix with* $\mathbf{u}_1, \ldots, \mathbf{u}_k$ *as its columns. Set* $\mathbf{B} := \mathbf{V}^T\mathbf{A}\mathbf{V} \in \mathbb{R}^{k\times k}$ *and arrange the eigenvalues of* $\mathbf{A}$ *and* $\mathbf{B}$ *as above. Then,* $\lambda_d^{\mathbf{A}} \leq \lambda_d^{\mathbf{B}}$ *for* $1 \leq d \leq k$*.*

Proof. If $k < n$, pick $n-k$ additional orthonormal vectors $\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n \in \mathbb{R}^n$, and let $\mathbf{U} \in \mathbb{R}^{n\times n}$ be an extension of $\mathbf{V}$ by adding these last orthonormal vectors as columns. Then, $\mathbf{U}$ is an orthogonal matrix, i.e. $\mathbf{U}^T\mathbf{U} = 1_n$. By lemma A.10, we see that $\mathbf{A}$ and $\mathbf{U}^T\mathbf{A}\mathbf{U}$ share the same eigenvalues.

To finish our argument, see that removing the last $n-k$ rows and columns of $\mathbf{U}^T\mathbf{A}\mathbf{U}$ gives us back $\mathbf{B}$. Hence, lemma 0.5 gives us

$$\lambda_d^{\mathbf{A}} = \lambda_d^{\mathbf{U}^T\mathbf{A}\mathbf{U}} \leq \lambda_d^{\mathbf{B}},$$

which is what we wanted to show. □

Corollary 0.7. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *as above. Then,* $\mathrm{Tr}(\mathbf{B}) \geq \sum_{d=1}^{k} \lambda_d^{\mathbf{A}}$*.*

Proof. Combining corollary 0.6 and the definition of Tr yields

$$\mathrm{Tr}(\mathbf{B}) = \sum_{d=1}^{k} \lambda_d^{\mathbf{B}} \geq \sum_{d=1}^{k} \lambda_d^{\mathbf{A}}.$$

□

For our proof of theorem 0.1, we will rewrite our original objective

$$\begin{aligned} &\underset{\mathbf{M}}{\text{minimize}} && \mathrm{Tr}(\mathbf{M}^T\mathbf{L}\mathbf{M}), \\ &\text{subject to} && \mathbf{M}^T\mathbf{D}\mathbf{M} = \mathbf{1}^{k\times k} \text{ and } \mathbf{M} \in \mathbb{R}^{n\times k}. \end{aligned}$$

to suit the form of corollary 0.6. To do so, let $\mathbf{L_N} := \mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}} \in \mathbb{R}^{n\times n}$ and $\mathbf{H} := \mathbf{D}^{\frac{1}{2}}\mathbf{M}$. The objective then becomes

$$\begin{aligned} &\underset{\mathbf{H}}{\text{minimize}} && \mathrm{Tr}(\mathbf{H}^T\mathbf{L_N}\mathbf{H}), \\ &\text{subject to} && \mathbf{H}^T\mathbf{H} = \mathbf{1}^{k\times k} \text{ and } \mathbf{H} \in \mathbb{R}^{n\times k}. \end{aligned}$$

To enhance the structure of the proof, we will show one property of $\mathbf{L_N}$ that we will use.

Lemma 0.8. *If $\mathbf{v}$ is an eigenvector of $\mathbf{L_N}$ with eigenvalue λ, then $\mathbf{u} := \mathbf{D}^{-\frac{1}{2}}\mathbf{v}$ is a solution to the generalized eigenvalue problem $\mathbf{L}\mathbf{u} = \lambda\mathbf{D}\mathbf{u}$.*

Proof.

$$\begin{aligned} \mathbf{L_N}\mathbf{v} = \lambda\mathbf{v} &\iff \mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}}\mathbf{v} = \lambda\mathbf{v} \\ &\iff \mathbf{L}\mathbf{D}^{-\frac{1}{2}}\mathbf{v} = \lambda\mathbf{D}^{\frac{1}{2}}\mathbf{v} && \text{(multiplying by } \mathbf{D}^{\frac{1}{2}} \text{ from the left)} \\ &\iff \mathbf{L}(\mathbf{D}^{-\frac{1}{2}}\mathbf{v}) = \lambda\mathbf{D}(\mathbf{D}^{-\frac{1}{2}}\mathbf{v}) && \text{(by rearranging)} \\ &\iff \mathbf{L}\mathbf{u} = \lambda\mathbf{D}\mathbf{u}. \end{aligned}$$

□

Proof of theorem 0.1. By definition, $\mathbf{L_N} \in \mathbb{R}^{n\times n}$ is a symmetric matrix. Hence, corollary 0.7 applies, so

$$\mathrm{Tr}(\mathbf{H}^T\mathbf{L_N}\mathbf{H}) \geq \sum_{d=1}^{k} \lambda_d^{\mathbf{L_N}},$$

for all $\mathbf{H} \in \mathbb{R}^{n\times k}$ with $\mathbf{H}^T\mathbf{H} = \mathbf{1}^{k\times k}$.

Now set $\mathbf{H}$ to be the matrix having the first k eigenvectors of $\mathbf{L_N}$, $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$, as its columns. It then follows that

$$\begin{aligned} \mathrm{Tr}(\mathbf{H}^T\mathbf{L_N}\mathbf{H}) &= \sum_{d=1}^{k} \mathbf{v}_d^T\mathbf{L_N}\mathbf{v}_d \\ &= \sum_{d=1}^{k} \mathbf{v}_d^T\lambda_d^{\mathbf{L_N}}\mathbf{v}_d \\ &= \sum_{d=1}^{k} (\mathbf{v}_d^T\mathbf{v}_d)\lambda_d^{\mathbf{L_N}} \\ &= \sum_{d=1}^{k} \lambda_d^{\mathbf{L_N}} && \text{(since } \mathbf{v}_d^T\mathbf{v}_d = 1\text{)} \end{aligned}$$

Combining everything, we get

$$\min_{\mathbf{H}\in\mathbb{R}^{n\times k}} \mathrm{Tr}(\mathbf{H}^T\mathbf{L_N}\mathbf{H}) = \sum_{d=1}^{k} \lambda_d^{\mathbf{L_N}},$$

and this minimum is reached by the matrix $\mathbf{H}$ having the first k eigenvectors of $\mathbf{L_N}$ as its columns. If we substitute back $\mathbf{M} = \mathbf{D}^{-\frac{1}{2}}\mathbf{H}$, we see by lemma 0.8 that the minimum for the original objective is reached by the matrix $\mathbf{M}$ having the first k eigenvectors of the generalized eigenvalue problem $\mathbf{L}\mathbf{u} = \lambda\mathbf{D}\mathbf{u}$, which is what we wanted to show. □

Bibliography

[1] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[2] Ahmed H Sameh and John A Wisniewski. A trace minimization algorithm for the generalized eigenvalue problem. *SIAM Journal on Numerical Analysis*, 19(6):1243–1259, 1982.

A Mathematical prerequisites

A.1 Linear algebra

A.1.1 Vector space

A vector space over $\mathbb{R}$ is a non-empty set V equipped with an addition and a (scalar) multiplication operation, defined

$$\begin{aligned} + : V \times V &\to V \\ (\mathbf{x}, \mathbf{y}) &\mapsto \mathbf{x} + \mathbf{y} \end{aligned}$$

and

$$\begin{aligned} \cdot : \mathbb{R} \times V &\to V \\ (\lambda, \mathbf{y}) &\mapsto \lambda \cdot \mathbf{y} = \lambda \mathbf{y}, \end{aligned}$$

such that the following conditions hold:

- For all $\mathbf{x}, \mathbf{y} \in V : \mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
- For all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V : (\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$
- For all $\mathbf{x}, \mathbf{y} \in V$ there exists a unique $\mathbf{z} \in V$ such that $\mathbf{x} + \mathbf{z} = \mathbf{y}$
- For all $\mathbf{x} \in V$ and $\lambda, \lambda' \in \mathbb{R}$: $(\lambda\lambda')\mathbf{x} = \lambda(\lambda'\mathbf{x})$
- For all $\mathbf{x} \in V$ and $\lambda, \lambda' \in \mathbb{R}$: $(\lambda + \lambda')\mathbf{x} = \lambda\mathbf{x} + \lambda'\mathbf{x}$
- For all $\mathbf{x}, \mathbf{y} \in V$ and $\lambda \in \mathbb{R}$: $\lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$
- For all $\mathbf{x} \in V$: $1\mathbf{x} = \mathbf{x}$.

Elements of the vector space V are called vectors. In this thesis, we will only use Euclidean vector spaces $\mathbb{R}^d$ with $d > 1$ with commonly defined dot product and norms. That is, for $\mathbf{x} := [x_1, \cdots, x_d]^T, \mathbf{y} := [y_1, \ldots, y_d]^T \in \mathbb{R}^d$, we denote the dot product between $\mathbf{x}$ and $\mathbf{y}$ by

$$\mathbf{x}^T\mathbf{y} = \begin{bmatrix} x_1 & \cdots & x_d \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_d \end{bmatrix} = \sum_{i=1}^{d} x_i y_i,$$

and the (Euclidean) norm of $\mathbf{x}$ by

$$||\mathbf{x}|| = \sqrt{\mathbf{x}^T\mathbf{x}}.$$

Lemma A.1.

$$||\mathbf{x} - \mathbf{y}||^2 = ||\mathbf{x}||^2 - 2\mathbf{x}^T\mathbf{y} + ||\mathbf{y}||^2.$$

Proof. We have

$$\begin{aligned} ||\mathbf{x} - \mathbf{y}||^2 &= (\mathbf{x} - \mathbf{y})^T(\mathbf{x} - \mathbf{y}) \\ &= \sum_{i=1}^{d}(x_i - y_i)^2 \\ &= \sum_{i=1}^{d}(x_i)^2 + \sum_{i=1}^{d}(y_i)^2 - 2\sum_{i=1}^{d} x_i y_i \\ &= ||\mathbf{x}||^2 - 2\mathbf{x}^T\mathbf{y} + ||\mathbf{y}||^2. \end{aligned}$$

□

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are called orthogonal if $\mathbf{x}^T\mathbf{y} = 0$, and orthonormal if they are orthogonal and additionally $||\mathbf{x}|| = ||\mathbf{y}|| = 1$.

A linear combination of vectors $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$ is a vector $\mathbf{x} \in \mathbb{R}^d$ such that

$$\mathbf{x} = \lambda_1\mathbf{x}_1 + \cdots + \lambda_n\mathbf{x}_n$$

for some $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. Vectors $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$ are said to be linearly independent if the linear combination such that

$$\lambda_1\mathbf{x}_1 + \cdots + \lambda_n\mathbf{x}_n = \mathbf{0}^d$$

is achieved only with $\lambda_1 = \cdots = \lambda_n = 0$. The span of vectors $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$, denoted $\mathrm{span}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is the set of all linear combinations of $\mathbf{x}_1, \dots, \mathbf{x}_n$.

Lemma A.2. *If $\mathbf{x}_1, \dots, \mathbf{x}_n$ are orthonormal, they are linearly independent.*

Proof. Let $\lambda_1\mathbf{x}_1 + \cdots + \lambda_n\mathbf{x}_n = \mathbf{0}^d$, and consider that for any $1 \le j \le n$, we have

$$\begin{aligned}\mathbf{x}_j^T \cdot (\lambda_1\mathbf{x}_1 + \cdots + \lambda_n\mathbf{x}_n) &= (\lambda_1\mathbf{x}_j^T\mathbf{x}_1 + \cdots + \lambda_n\mathbf{x}_j^T\mathbf{x}_n) \\ &= \lambda_j\mathbf{x}_j^T\mathbf{x}_j.\end{aligned}$$

Since $\mathbf{x}_j^T\mathbf{x}_j = \mathbf{0}^d$ implies $\mathbf{x}_j^T$ (which contradicts orthonormality) it holds that $\lambda_j = 0$. Hence, $\lambda_1 = \cdots = \lambda_n = 0$. □

A subspace $W \subset V$ is a subset of V such that

- $\mathbf{0} \in W$
- For all $\mathbf{x}, \mathbf{y} \in W$: $\mathbf{x} + \mathbf{y} \in W$, and
- For all $\mathbf{x} \in W$, $\lambda \in \mathbb{R}$: $\lambda\mathbf{x} \in W$.

Given two subspaces $W, W' \subset V$, the sets $W \cap W'$ and $W + W' := \{\mathbf{w} + \mathbf{w}' \mid \mathbf{w} \in W, \mathbf{w}' \in W'\}$ are also subspaces of V. The dimension of a subspace W – denoted $\dim(W)$ – is the largest number k such that there exist k linearly independent vectors in W.

Lemma A.3. *Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be orthonormal. Then it holds that $\dim(\mathrm{span}(\mathbf{x}_1, \dots, \mathbf{x}_n)) = n$.*

Proof. By lemma A.2 we know $\dim(\mathrm{span}(\mathbf{x}_1, \dots, \mathbf{x}_n)) \ge n$. By definition, all elements $\mathbf{x}$ of $\mathrm{span}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ are linear combinations of $\mathbf{x}_1, \dots, \mathbf{x}_n$, and thus any $\mathbf{x} \in \mathrm{span}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ can be written as $\mathbf{x} = \lambda_1\mathbf{x}_1 + \cdots + \lambda_n\mathbf{x}_n$, or $\mathbf{0}^d = \lambda_1\mathbf{x}_1 + \cdots + \lambda_n\mathbf{x}_n - \mathbf{x}$. Since at least the last scalar is non-zero, $\mathbf{x}$ is not linearly independent. Hence, $\dim(\mathrm{span}(\mathbf{x}_1, \dots, \mathbf{x}_n)) = n$. □

Lemma A.4. *For two subspaces $W, W' \subseteq V$ of a vector space V, we have $\dim(W) + \dim(W') = \dim(W + W')$ if $W \cap W' = \{\mathbf{0}^d\}$.*

Proof. Let $\dim(W) := r$ and $\dim(W') := s$, and let $\mathbf{u}_1, \dots, \mathbf{u}_r \in W$, $\mathbf{v}_1, \dots, \mathbf{v}_s \in W'$ be two linearly independent sets of vectors. We will show that $\mathbf{u}_1, \dots, \mathbf{u}_r, \mathbf{v}_1, \dots, \mathbf{v}_s$ are all linear independent from each other. To do so, let

$$\mathbf{0}^d = \sum_{i=1}^{r} \lambda_i\mathbf{u}_i + \sum_{j=1}^{s} \lambda'_j\mathbf{v}_j,$$

or

$$\sum_{i=1}^{r} \lambda_i\mathbf{u}_i = -\sum_{j=1}^{s} \lambda'_j\mathbf{v}_j.$$

Let $\mathbf{x} := \sum_{i=1}^r \lambda_i\mathbf{u}_i = -\sum_{j=1}^s \lambda'_j\mathbf{v}_j$. By definition of subspaces, $\mathbf{x} \in W$ and also $\mathbf{x} \in W'$, so $\mathbf{x} \in W \cap W'$. Since $W \cap W' = \{\mathbf{0}^d\}$ however, we know, $\sum_{i=1}^r \lambda_i\mathbf{u}_i = -\sum_{j=1}^s \lambda'_j\mathbf{v}_j = \mathbf{0}^d$. Since $\mathbf{u}_1, \dots, \mathbf{u}_r$ and $\mathbf{v}_1, \dots, \mathbf{v}_s$ are linearly independent, this means $\lambda_1 = \cdots = \lambda_r = \lambda'_1 = \cdots = \lambda'_s = 0$, which in turn means $\mathbf{u}_1, \dots, \mathbf{u}_r, \mathbf{v}_1, \dots, \mathbf{v}_s$ are linearly independent.

Now let $\mathbf{x} \in W + W'$, such that it can be written as $\mathbf{x} = \sum_{i=1}^r \lambda_i\mathbf{u}_i + \sum_{j=1}^s \lambda'_j\mathbf{v}_j$. By the same argument as lemma A.3 we see $\mathbf{x}$ is not linearly independent from $\mathbf{u}_1, \dots, \mathbf{u}_r, \mathbf{v}_1, \dots, \mathbf{v}_s$. Since $\mathbf{x}$ was chosen arbitrarily, there exist at most $r+s$ linearly independent vectors in $W+W'$, i.e. $\dim(W+W')=r+s=\dim(W)+\dim(W')$, which is what we wanted to show. $\square$

Lemma A.5. *$dim(\mathbb{R}^d)=d$.*

Proof. Clearly, vectors $[1,0\cdots 0]^T,[0,1\cdots 0]^T,\ldots,[0,0\cdots 1]^T \in \mathbb{R}^d$ are linearly independent. Therefore, $\dim(\mathbb{R}^d)\geq d$. Now let $\mathbf{x}=[x_1\cdots x_d]^T$ be some other element of $\mathbb{R}^d$. Then, $\mathbf{x}$ can be written as

$$\mathbf{x}=x_1\cdot[10\cdots 0]^T+\cdots+x_d\cdot[00\cdots 1]^T,$$

which means $\mathbf{x}$ is not linearly independent from $[10\cdots 0]^T,[01\cdots 0]^T,\ldots,[00\cdots 1]^T$. Since $\mathbf{x}$ was chosen arbitrarily, $\dim(\mathbb{R}^d)=d$. $\square$

A.1.2 Matrices

A matrix is a rectangular array of objects – which, in the context of this thesis, are all real numbers – called entries. A matrix with n rows and k columns is a $n\times k$ matrix, and the set of all such matrices is denoted $\mathbb{R}^{n\times k}$ (n and k are also called the dimensions). The entry in the i-th row and j-th column of a matrix $\mathbf{M}\in\mathbb{R}^{n\times k}$ is denoted $(\mathbf{M})_{ij}$, such that the matrix can be portrayed visually as

$$\mathbf{M}=\begin{bmatrix}(\mathbf{M})_{11} & (\mathbf{M})12 & \ldots & (\mathbf{M})1k\\ (\mathbf{M})21 & (\mathbf{M})22 & \ldots & (\mathbf{M})2k\\ \vdots & \vdots & \ddots & \vdots\\ (\mathbf{M})n1 & (\mathbf{M})n2 & \ldots & (\mathbf{M})nk\end{bmatrix}.$$

A d-dimensional (row) vector is also a $1\times d$-dimensional matrix. There are a couple basic operations defined on matrices: addition, (scalar) multiplication and transposition. Addition and scalar multiplication function entrywise, i.e. for $\mathbf{M},\mathbf{M}'\in\mathbb{R}^{n\times k}$, $\lambda\in\mathbb{R}$:

$$(\mathbf{M}+\mathbf{M}')_{ij}=(\mathbf{M})_{ij}+(\mathbf{M}')_{ij},$$
$$(\lambda\mathbf{M})_{ij}=\lambda(\mathbf{M})_{ij}.$$

Switching rows and columns, we get the transpose of $\mathbf{M}$, denoted $\mathbf{M}^T$, defined as

$$(\mathbf{M}^T)_{ij}=(\mathbf{M})_{ji}.$$

Multiplication of two matrices is defined for matrices $\mathbf{M}\in\mathbb{R}^{n\times k}$ and $\mathbf{M}'\in\mathbb{R}^{k\times n'}$ – such that the number of columns in $\mathbf{M}$ equals the number of rows in $\mathbf{M}'$ – and returns a matrix $\mathbf{M}\mathbf{M}'\in\mathbb{R}^{n\times n'}$ such that

$$(\mathbf{M}\mathbf{M}')_{ij}=\sum_{r=1}^{k}(\mathbf{M})_{ir}(\mathbf{M}')_{rj}=\mathbf{m}_i^T\mathbf{m}'_j,$$

where $\mathbf{m}_i$ and $\mathbf{m}'_j$ refer to the i-th row of $\mathbf{M}$ and j-th column of $\mathbf{M}'$ respectively. A few properties using and connecting the operations include:

- $\mathbf{M}(\mathbf{M}'+\mathbf{M}'')=\mathbf{M}\mathbf{M}'+\mathbf{M}\mathbf{M}''$
- $\mathbf{M}(\mathbf{M}'\mathbf{M}'')=(\mathbf{M}\mathbf{M}')\mathbf{M}''$
- $(\mathbf{M}^T)^T=\mathbf{M}$
- $(\mathbf{M}+\mathbf{M}')^T=\mathbf{M}^T+\mathbf{M}'^T$

If $\mathbf{M}^T=\mathbf{M}$ we call $\mathbf{M}$ symmetric. A matrix $\mathbf{M}\in\mathbb{R}^{n\times n}$ is called orthogonal if the columns of $\mathbf{M}$ form a set of orthonormal vectors, such that $\mathbf{M}^T\mathbf{M}=\mathbf{1}^{n\times n}$. Here, $\mathbf{1}^{n\times n}$ is the identity matrix defined as $(\mathbf{1})_{ij}=\delta_{ij}$ where δ_{ij} is the Kronecker delta. The $n\times k$ matrix with only zero entries is denoted $\mathbf{0}^{n\times k}$.

Multiplication of a matrix with a vectors induces another important notion we will use: eigenvectors. For any square matrix $\mathbf{M}\in mathbbR^{n\times n}$, $\mathbf{x}\in mathbbR^n$, we call $\mathbf{x}$ an eigenvector of $\mathbf{M}$ if

$$\mathbf{M}\mathbf{x}=\lambda\mathbf{x}$$

for some $\lambda\in\mathbb{R}$, and call λ a eigenvalue of $\mathbf{M}$.

Lemma A.6. *If* $\mathbf{v}$ *is an eigenvector of* $\mathbf{M}$ *with eigenvalue* λ *than* $\alpha\mathbf{v}$ *is also an eigenvector* $\mathbf{M}$ *with eigenvalue* λ *for any* $0 \neq \alpha \in \mathbb{R}$.

Proof. Let λ be the eigenvalue corresponding to $\mathbf{v}$. Then, we have

$$\mathbf{M}(\alpha\mathbf{v} = \alpha(\mathbf{M}\mathbf{v}) = \alpha\lambda\mathbf{v} = \lambda(\alpha\mathbf{v}).$$

□

Lemma A.6 tells us it is silly to speak of *the* eigenvectors of $\mathbf{M}$, when in fact there are an infinite amount of them corresponding to every distinct eigenvalue. In particular, for every λ we can construct an eigenvector of unit length by picking some eigenvector $\mathbf{v}$ and determining

$$\frac{1}{||\mathbf{v}||}\mathbf{v}.$$

Lemma A.7. *Let* $\mathbf{m}$ *be symmetric. Then, any two eigenvectors of* $\mathbf{M}$ *with distinct eigenvectors are orthogonal.*

Proof. Let $\mathbf{x}, \mathbf{y}$ be distinct eigenvectors with eigenvalues λ and λ', $\lambda \neq \lambda'$. Then by definition the following holds:

$$\begin{aligned}\mathbf{M}\mathbf{x} &= \lambda\mathbf{x}\\ \mathbf{M}\mathbf{y} &= \lambda'\mathbf{y}.\end{aligned}$$

It now follows that

$$\begin{aligned}\lambda(\mathbf{x}^T\mathbf{y}) &= (\lambda\mathbf{x})^T\mathbf{y}\\ &= (\mathbf{M}\mathbf{x})^T\mathbf{y}\\ &= \mathbf{x}^T\mathbf{M}^T\mathbf{y}\\ &= \mathbf{x}^T\mathbf{M}\mathbf{y}\\ &= \mathbf{x}^T\lambda'\mathbf{y}\\ &= \lambda'(\mathbf{x}^T\mathbf{y}),\end{aligned}$$

meaning $0 = (\lambda - \lambda')(\mathbf{x}^T\mathbf{y})$. Since $\lambda \neq \lambda'$, we have $\mathbf{x}^T\mathbf{y} = 0$, which means $\mathbf{x}$ and $\mathbf{y}$ are orthogonal. □

A $n \times n$ symmetric matrix $\mathbf{M}$ is called positive semi-definite if and only if $\mathbf{x}^T\mathbf{M}\mathbf{x} \geq \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

Lemma A.8. *If* $\mathbf{M}$ *is positive semi-definite, it has only non-negative eigenvalues.*

Proof. Let λ be some eigenvalue of $\mathbf{M}$ and let $\mathbf{v}$ be some corresponding eigenvector. Then it holds that $0 \leq \mathbf{v}^T\mathbf{M}\mathbf{v} = \mathbf{v}^T(\lambda\mathbf{v}) = (\mathbf{v}^T\mathbf{v})\lambda$. Since $\mathbf{v}^T\mathbf{v} \geq 0$ (easily checked) it must be the case that $\lambda \geq 0$. □

Combining lemma A.6, A.7 and A.8, we see that for any positive semi-definite matrix $\mathbf{M}$, we know $\mathbf{M}$ has eigenvalues $0 \leq \lambda_1 \leq \cdots \leq \lambda_k$ (for some $k > 0$), such that there exists an orthonormal set of eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ corresponding to those eigenvalues.

For any eigenvalue λ, we define its eigenspace $E_{\mathbf{M}}(\lambda)$ as the set of all eigenvectors corresponding to this eigenvalue, i.e. for some $\mathbf{M} \in mathbbR^{n \times n}$ it is defined as

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{M}\mathbf{x} = \lambda\mathbf{x}\}.$$

Lemma A.9. *For any* $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y} \in \mathbb{R}^d$*, with* $\mathbf{x}_i := \left[x_i^1, \ldots, x_i^d\right]^T$ *for* $1 \leq i \leq n$*, we have*

$$\sum_{i=1}^{n} \mathbf{x}_i^T\mathbf{y} = (\sum_{i=1}^{n} \mathbf{x}_i)^T\mathbf{y}$$

Proof.

$$\sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{y} = (\sum_{i=1}^{n} \mathbf{x}_i^T) \cdot \mathbf{y}$$
$$= (\sum_{i=1}^{n} \mathbf{x}_i)^T \cdot \mathbf{y}. \qquad \text{(since } (\mathbf{M}+\mathbf{M}')^T = \mathbf{M}^T + \mathbf{M}'^T\text{)}$$

□

Lemma A.10. *Let* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$. *Then* $\lambda \neq 0$ *is an eigenvalue of* $\mathbf{AB}$ *iff* λ *is an eigenvalue of* $\mathbf{BA}$.

Proof. Let $\mathbf{u}$ be an eigenvector of $\mathbf{AB}$ associated with eigenvalue λ, i.e. $\mathbf{ABu} = \lambda\mathbf{u}$. Let $\mathbf{v} := \mathbf{Bu}$. Then,

$$\begin{aligned}\mathbf{BAv} &= \mathbf{BABu}\\ &= \mathbf{B}(\mathbf{ABu})\\ &= \mathbf{B}\lambda\mathbf{u}\\ &= \lambda(\mathbf{Bu})\\ &= \lambda\mathbf{v}.\end{aligned}$$

Hence, λ is an eigenvalue of BA too. The other implication follows analogously, mutatis mutandis. □

Lastly, the trace of a square matrix $\mathbf{M}$, denoted $\mathrm{Tr}(\mathbf{M})$ is the sum of its diagonal elements, i.e.

$$\mathrm{Tr}(\mathbf{M}) = \sum_{i=1}^{n} (\mathbf{M})_{ii}.$$

It can be shown that the trace is equal to the sum of the eigenvalues of $\mathbf{M}$.