Prerequisites from Matrix Algebra

In this chapter we review some basic results from matrix algebra.
Real $m \times n$ matrices will be written as:

$$\mathbf{M} = (m_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} \in \mathbb{R}^{m \times n}$$

Diagonal matrices are written as $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^{n \times n}$. A Matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$ is called orthogonal if :

$$\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}_n$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix. $\mathcal{O}(n)$ denotes the set of orthogonal $n \times n$ matrices.

Theorem 2.1 (Singular Value Decomposition, SVD)**.** *Given* $\mathbf{M} \in \mathbb{R}^{m \times n}$, *there exists* $\mathbf{U} \in \mathcal{O}(m)$ *and* $\mathbf{V} \in \mathcal{O}(n)$ *and some* $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ *with non-negative entries in its diagonal and zeros otherwise such that:*

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T.$$

The diagonal elements of $\mathbf{\Sigma}$ *are called singular values. The columns of* $\mathbf{U}$ *and* $\mathbf{V}$ *are called left and right singular vectors of* $\mathbf{M}$.

Remark 1. If $m < n$, say, SVD may be written as :

$$\exists \mathbf{U} \in \mathbb{R}^{m \times n}, \mathbf{U}\mathbf{U}^T = \mathbf{I}_m, \exists \mathbf{V} \in \mathcal{O}(n); \exists \mathbf{\Sigma} \in \mathbb{R}^{n \times n} \text{ diagonal, such that: } \mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T.$$

Theorem 2.2 (Spectral Decomposition)**.** *Given* $\mathbf{M} \in \mathbb{R}^{n \times n}$ *symmetric, there exists* $\mathbf{V} \in \mathcal{O}(n)$, $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ *and* $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_n)$ *such that:*

$$\mathbf{M} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T.$$

$\mathbf{v}_i$*'s are eigenvectors of* $\mathbf{M}$ *with the eigenvalues* λ_i.

- If for the symmetric matrix $\mathbf{M}$, $\lambda_i > 0$, $i = 1, \ldots, n$, then $\mathbf{M}$ is called positive definite (p.d.) and writes as $\mathbf{M} \succ 0$.
  If for the symmetric matrix $\mathbf{M}$, $\lambda_i \geq 0$, $i = 1, \ldots, n$, then $\mathbf{M}$ is called non-negative definite (n.n.d.) and writes as $\mathbf{M} \succeq 0$.

- If $\mathbf{M}$ is non-negative definite, then it has a Cholesky decomposition

$$\mathbf{M} = \mathbf{V}\mathbf{\Lambda}^{\frac{1}{2}}(\mathbf{V}\mathbf{\Lambda}^{\frac{1}{2}})^T,$$

  where $\mathbf{\Lambda}^{\frac{1}{2}} = \text{diag}(\lambda_1^{\frac{1}{2}}, \ldots, \lambda_n^{\frac{1}{2}})$.

- $\mathbf{M} \succeq 0 \iff \mathbf{x}^T\mathbf{M}\mathbf{x} \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n$
- $\mathbf{M} \succ 0 \iff \mathbf{x}^T\mathbf{M}\mathbf{x} > 0 \quad \forall \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{x} \neq 0.$

Definition 2.3. (a) Given $\mathbf{M} = (m_{ij}) \in \mathbb{R}^n$, $\operatorname{tr}(\mathbf{M}) = \sum_{i=1}^n m_{ii}$ is called the trace of $\mathbf{M}$.

(b) Given $\mathbf{M} \in \mathbb{R}^{n\times n}$, $\|\mathbf{M}\|_F = \sqrt{\sum_{i,j} m_{ij}^2} = \sqrt{\operatorname{tr}(\mathbf{M}^T\mathbf{M})}$ is called the Frobenius norm.

(c) Given $\mathbf{M} \in \mathbb{R}^{n\times n}$, $\mathbf{M}$ symmetric, $\|\mathbf{M}\|_S = \max_{1\leq i\leq n} |\lambda_i|$ is called the spectral norm.

- It holds that $\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$, $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times m}$.
- $\operatorname{tr}(\mathbf{M}) = \sum_{i=1}^n \lambda_i(\mathbf{M})$, $\det(\mathbf{M}) = \prod_{i=1}^n \lambda_i(\mathbf{M})$.

  A simple proof of this statement uses the spectral decomposition of $\mathbf{M}$, for symmetric $\mathbf{M}$. First using the invariance property of trace under matrix commutation, we have

$$\begin{aligned}
\operatorname{tr}(\mathbf{M}) &= \operatorname{tr}(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T) \\
&= \operatorname{tr}(\boldsymbol{\Lambda}\mathbf{V}^T\mathbf{V}) \\
&\overset{(a)}{=} \operatorname{tr}(\boldsymbol{\Lambda}\mathbf{I}_n) \\
&= \operatorname{tr}(\boldsymbol{\Lambda}) = \sum_{i=1}^n \lambda_i(\mathbf{M}),
\end{aligned}$$

  where (a) follows from the fact that $\mathbf{V}$ is an orthogonal matrix. Similarly, the spectral decomposition of symmetric matrix $\mathbf{M}$ can be used to prove the respective statement for the determinant.

$$\begin{aligned}
\det(\mathbf{M}) &= \det(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T) \\
&= \det(\boldsymbol{\Lambda})\det(\mathbf{V}^T)\det(\mathbf{V}) \\
&\overset{(b)}{=} \det(\boldsymbol{\Lambda}) = \prod_{i=1}^n \lambda_i(\mathbf{M}),
\end{aligned}$$

  where (b) follows from the fact that the determinant of an orthogonal matrix is either $+1$ or -1.

Theorem 2.4 (Ky Fan, 1950 ([Fan50])). *Let $\mathbf{M} \in \mathbb{R}^{n\times n}$ be a symmetric matrix with eigenvalues $\lambda_1(\mathbf{M}) \geq \cdots \geq \lambda_n(\mathbf{M})$ and let $k \leq n$. We have:*

$$\max_{\mathbf{V}\in\mathbb{R}^{n\times k}, \mathbf{V}^T\mathbf{V}=\mathbf{I}_k} \operatorname{tr}(\mathbf{V}^T\mathbf{M}\mathbf{V}) = \sum_{i=1}^k \lambda_i(\mathbf{M}).$$

and

$$\min_{\mathbf{V}\in\mathbb{R}^{n\times k}, \mathbf{V}^T\mathbf{V}=\mathbf{I}_k} \operatorname{tr}(\mathbf{V}^T\mathbf{M}\mathbf{V}) = \sum_{i=1}^k \lambda_{n-i+1}(\mathbf{M}).$$

Proof. First of all, see that $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_k]$ where $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ are orthonormal vectors. If the span of orthonormal vectors $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ is same as the span of $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$, then for $\mathbf{W} = [\mathbf{w}_1 \dots \mathbf{w}_k]$, there is a unitary matrix $\mathbf{A} \in \mathbb{R}^{k\times k}$, which is a basis changing matrix, such that $\mathbf{V} = \mathbf{WA}$. We have:

$$\mathrm{tr}(\mathbf{V}^T\mathbf{MV}) = \mathrm{tr}(\mathbf{A}^T\mathbf{W}^T\mathbf{MWA}) = \mathrm{tr}(\mathbf{W}^T\mathbf{MWAA}^T) = \mathrm{tr}(\mathbf{W}^T\mathbf{MW}).$$

Also note that since $\mathrm{tr}(\mathbf{V}^T\mathbf{MV}) = \sum_{i=1}^k \mathbf{v}_i^T\mathbf{M}\mathbf{v}_i$:

$$\sum_{i=1}^k \mathbf{v}_i^T\mathbf{M}\mathbf{v}_i = \sum_{i=1}^k \mathbf{w}_i^T\mathbf{M}\mathbf{w}_i.$$

The proof follows an iterative procedure. Suppose that $\mathbf{u}_1, \dots, \mathbf{u}_n$ are eigenvectors of $\mathbf{M}$ corresponding to $\lambda_1(\mathbf{M}) \geq \cdots \geq \lambda_n(\mathbf{M})$ and $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$. For $k = 1$, every vector $\mathbf{v}$ can be written as $\mathbf{v} = a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n = \mathbf{Ua}$, where $\mathbf{a} = [a_1 \dots a_n]^T$. We have:

$$\max_{\mathbf{v}\in\mathbb{R}^n, \mathbf{v}^T\mathbf{v}=1} \mathbf{v}^T\mathbf{Mv} = \max_{\mathbf{a}\in\mathbb{R}^n, \mathbf{a}^T\mathbf{a}=1} \mathbf{a}^T\mathbf{U}^T\mathbf{MUa} = \max_{\mathbf{a}\in\mathbb{R}^n, \mathbf{a}^T\mathbf{a}=1} \sum_{i=1}^n \lambda_i(\mathbf{M})a_i^2.$$

Therefore for $k = 1$, $\max_{\mathbf{v}\in\mathbb{R}^n, \mathbf{v}^T\mathbf{v}=1} \mathbf{v}^T\mathbf{Mv} = \lambda_1(\mathbf{M})$.

For $k = 2$, see that for each orthonormal vector $\{\mathbf{v}_1, \mathbf{v}_2\}$, one can find two orthonormal vectors $\{\mathbf{v}_1^*, \mathbf{v}_2^*\}$ with the same span so that $\mathbf{v}_2^*$ is inside the span of $\mathbf{u}_2, \dots, \mathbf{u}_n$. First of all, it can be seen that:

$$\sum_{i=1}^2 \mathbf{v}_i^T\mathbf{M}\mathbf{v}_i = \sum_{i=1}^2 (\mathbf{v}_i^*)^T\mathbf{M}\mathbf{v}_i^*.$$

Since $\mathbf{v}_2^*$ is inside the span of $\mathbf{u}_2, \dots, \mathbf{u}_n$, it can be written as $\mathbf{v}^* = a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n$, where $\mathbf{a} = [a_2 \dots a_n]^T$. It can be seen that:

$$(\mathbf{v}_2^*)^T\mathbf{M}\mathbf{v}_2^* = \sum_{i=2}^n \lambda_i(\mathbf{M})a_i^2 \leq \lambda_2(\mathbf{M}).$$

Moreover from the previous step $(\mathbf{v}_1^*)^T\mathbf{M}\mathbf{v}_1^* \leq \lambda_1(\mathbf{M})$. Hence:

$$\sum_{i=1}^2 \mathbf{v}_i^T\mathbf{M}\mathbf{v}_i \leq \lambda_1(\mathbf{M}) + \lambda_2(\mathbf{M}).$$

The upper bound is achievable by choosing $\mathbf{v}_1 = \mathbf{u}_1$ and $\mathbf{v}_2 = \mathbf{u}_2$. The procedure goes on iteratively. For a given k, the space spanned by $\mathbf{v}_1, \dots, \mathbf{v}_k$ has a $(k-1)$-dimensional subspace in intersection with the span of $\mathbf{u}_2, \dots, \mathbf{u}_n$. Find an orthonormal basis for this subspace $\mathbf{v}_2^*, \dots, \mathbf{v}_k^*$ and extend it with another $\mathbf{v}_1^*$ to an orthonormal basis for the space spanned by $\mathbf{v}_1, \dots, \mathbf{v}_k$. Then the sum $\sum_{i=2}^k (\mathbf{v}_i^*)^T\mathbf{M}\mathbf{v}_i^*$ is at most $\lambda_2(\mathbf{M}) + \cdots + \lambda_k(\mathbf{M})$ and $\mathbf{v}_1^{*T}\mathbf{M}\mathbf{v}_1^* \leq \lambda_1(\mathbf{M})$. Therefore:

$$\sum_{i=1}^k \mathbf{v}_i^T\mathbf{M}\mathbf{v}_i \leq \lambda_1(\mathbf{M}) + \cdots + \lambda_k(\mathbf{M}),$$

where the upper bound is achievable using $\mathbf{v}_i = \mathbf{u}_i$ for $i = 1, \dots, k$. $\square$

The special case of above statements for $k = 1$ writes as:

$$\max_{\|\mathbf{v}\|=1, \mathbf{v}\in\mathbb{R}^n} \mathbf{v}^T\mathbf{M}\mathbf{v} = \lambda_{\max}(\mathbf{M})$$

$$\min_{\|\mathbf{v}\|=1, \mathbf{v}\in\mathbb{R}^n} \mathbf{v}^T\mathbf{M}\mathbf{v} = \lambda_{\min}(\mathbf{M}).$$

Note that:

$$\max_{\|\mathbf{v}\|=1, \mathbf{v}\in\mathbb{R}^n} \mathbf{v}^T\mathbf{M}\mathbf{v} = \max_{\mathbf{v}\neq 0\in\mathbb{R}^n} \frac{\mathbf{v}^T\mathbf{M}\mathbf{v}}{\mathbf{v}^T\mathbf{v}}.$$

Theorem 2.5. *Given $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$, symmetric with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ and $\mu_1 \geq \cdots \geq \mu_n$, respectively. Then*

$$\sum_{i=1}^{n} \lambda_i \mu_{n-i+1} \leq \operatorname{tr}(\mathbf{A}\mathbf{B}) \leq \sum_{i=1}^{n} \lambda_i \mu_i.$$

Let $\lambda^+ = \max\{\lambda, 0\}$ denote the positive part of $\lambda \in \mathbb{R}$.

Theorem 2.6. *Given $\mathbf{M} \in \mathbb{R}^{n\times n}$ symmetric with spectral decomposition $\mathbf{M} = \mathbf{V}\operatorname{diag}(\lambda_1, \dots, \lambda_n)\mathbf{V}^T$, $\lambda_1 \geq \cdots \geq \lambda_n$. Then*

$$\min_{\mathbf{A}\succeq 0, \operatorname{rk}(\mathbf{A})\leq k} \|\mathbf{M} - \mathbf{A}\|_F^2$$

is attained at $\mathbf{A}^* = \mathbf{V}\operatorname{diag}(\lambda_1^+, \dots, \lambda_k^+, 0, \dots, 0)\mathbf{V}^T$ with optimum value $\sum_{i=1}^{k}(\lambda_i - \lambda_i^+)^2 + \sum_{i=k+1}^{n} \lambda_i^2$.

Proof.

$$\begin{aligned}
\|\mathbf{M} - \mathbf{A}\|^2 &= \|\mathbf{M}\|^2 - 2\operatorname{tr}(\mathbf{M}\mathbf{A}) + \|\mathbf{A}\|^2 \\
&\geq \sum_{i=1}^{n} \lambda_i^2 - 2\sum_{i=1}^{n} \lambda_i\mu_i + \sum_{i=1}^{n} \mu_i^2 \\
&= \sum_{i=1}^{n} (\lambda_i - \mu_i)^2 \\
&= \sum_{i=1}^{k} (\lambda_i - \mu_i)^2 + \sum_{i=k+1}^{n} (\lambda_i - 0)^2 \\
&\geq \sum_{i=1}^{k} (\lambda_i - \lambda_i^+)^2 + \sum_{i=k+1}^{n} \lambda_i^2.
\end{aligned}$$

Lower bound is attained if $\mathbf{A} = \mathbf{V}\operatorname{diag}(\lambda_1^+, \dots, \lambda_k^+, 0, \dots, 0)\mathbf{V}^T$. $\square$

Definition 2.7. (Löwner semi-ordering) Given $\mathbf{V}, \mathbf{W} \succeq 0$. Define $\mathbf{V} \preceq \mathbf{W}$ if and only if $\mathbf{W} - \mathbf{V} \succeq 0$. It can be shown that the relation "$\preceq$" imposes a semi-ordering on the set of non-negative definite matrices, i.e., it satisfies the following properties

- (reflexive) $\mathbf{V} \preceq \mathbf{V}$
- (anti-symmetric) $\mathbf{V} \preceq \mathbf{W}$ and $\mathbf{W} \preceq \mathbf{V} \implies \mathbf{V} = \mathbf{W}$
- (transitive) $\mathbf{U} \preceq \mathbf{V}$ and $\mathbf{V} \preceq \mathbf{W} \implies \mathbf{U} \preceq \mathbf{W}$.

Theorem 2.8. *Let* $\mathbf{V}$ *and* $\mathbf{W}$ *be two* $n \times n$ *non-negative definite matrices, such that* $\mathbf{V} = (v_{ij}) \preceq \mathbf{W} = (w_{ij})$*, with the eignevalues as:*

- $\lambda_1(\mathbf{V}) \geq \cdots \geq \lambda_n(\mathbf{V})$,
- $\lambda_1(\mathbf{W}) \geq \cdots \geq \lambda_n(\mathbf{W})$

(a) $\lambda_i(\mathbf{V}) \leq \lambda_i(\mathbf{W})$*, for* $i = 1, \ldots, n$

(b) $v_{ii} \leq w_{ii}$*, for* $i = 1, \ldots, n$

(c) $v_{ii} + v_{jj} - 2v_{ij} \leq w_{ii} + w_{jj} - 2w_{ij}$

(d) $\operatorname{tr}(\mathbf{V}) \leq \operatorname{tr}(\mathbf{W})$

(e) $\det(\mathbf{V}) \leq \det(\mathbf{W})$

Proof. Exercise. □

Projection and Isometry

Definition 2.9. The matrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is called a projection matrix, or idempotent, if $\mathbf{Q}^2 = \mathbf{Q}$. It is additionally called an orthogonal projection if additionally $\mathbf{Q}^T = \mathbf{Q}$.

The linear transformation $\mathbf{Q}$ maps onto $\operatorname{Im}(\mathbf{Q})$, a $k-$dimensional subspace of $\mathbb{R}^n$. Let $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{y} = \mathbf{Q}\mathbf{x} \in \operatorname{Im}(\mathbf{Q})$. Since $\mathbf{Q}$ is the projection matrix, $\mathbf{Q}\mathbf{y} = \mathbf{y}$. For an orthogonal projection, $\mathbf{x} - \mathbf{Q}\mathbf{x}$ is orthogonal to all vectors $\mathbf{y}$ in $\operatorname{Im}(\mathbf{Q})$ for every $\mathbf{x} \in \mathbb{R}^n$. To see this, note that there is a vector $\mathbf{z} \in \mathbb{R}^n$ such that $\mathbf{y} = \mathbf{Q}\mathbf{z}$. Then we have:

$$\mathbf{y}^T(\mathbf{x} - \mathbf{Q}\mathbf{x}) = \mathbf{z}^T\mathbf{Q}^T(\mathbf{x} - \mathbf{Q}\mathbf{x}).$$

Since for an orthogonal projection $\mathbf{Q}^T = \mathbf{Q}$ then:

$$\mathbf{z}^T\mathbf{Q}^T(\mathbf{x} - \mathbf{Q}\mathbf{x}) = \mathbf{z}^T\mathbf{Q}(\mathbf{x} - \mathbf{Q}\mathbf{x}) = \mathbf{z}^T(\mathbf{Q}\mathbf{x} - \mathbf{Q}^2\mathbf{x}) = \mathbf{z}^T(\mathbf{Q}\mathbf{x} - \mathbf{Q}\mathbf{x}) = 0.$$

Therefore $\mathbf{y}^T(\mathbf{x} - \mathbf{Q}\mathbf{x}) = 0$ and $\mathbf{x} - \mathbf{Q}\mathbf{x}$ is orthogonal to $\mathbf{y}$.

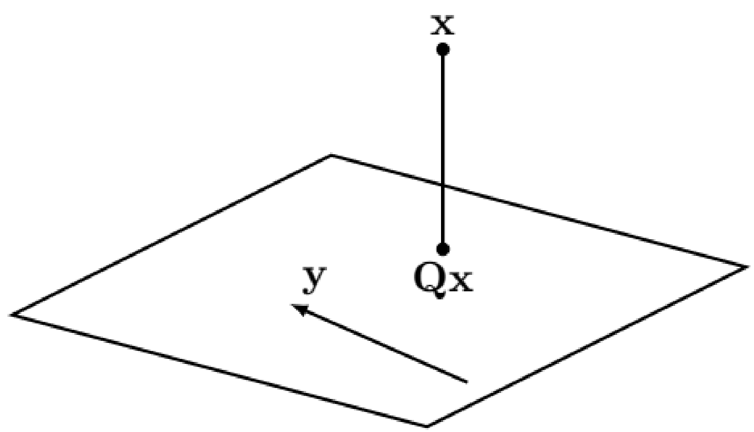

Figure 2.1: Orthogonal Projection

Lemma 2.10. *Let* $\mathbf{M} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ *be the spectral decomposition of* $\mathbf{M} \in \mathbb{R}^{n\times n}$ *and symmetric. For* $k \leq n$*, the matrix* $\mathbf{Q} = \sum_{i=1}^{k} \mathbf{v}_i\mathbf{v}_i^T$ *is an orthogonal projection onto* $\mathrm{Im}(\mathbf{Q}) = \langle \mathbf{v}_1, \ldots, \mathbf{v}_k \rangle$.

Proof. For $\mathbf{x} \in \mathbb{R}^n$, we have:

$$\mathbf{Qx} = \sum_{i=1}^{k} \mathbf{v}_i\mathbf{v}_i^T x = \sum_{i=1}^{k} (\mathbf{v}_i^T\mathbf{x})\mathbf{v}_i = \sum_{i=1}^{k} \gamma_i \mathbf{v}_i \in \mathrm{Im}(\mathbf{Q}).$$

Moreover:

$$\mathbf{Q}^2 = (\sum_{i=1}^{k} \mathbf{v}_i\mathbf{v}_i^T)(\sum_{i=1}^{k} \mathbf{v}_i\mathbf{v}_i^T) = \sum_{i=1}^{k} \mathbf{v}_i\mathbf{v}_i^T = \mathbf{Q}.$$

Finally $\mathbf{Q}$ is symmetric and therefore it is an orthogonal projection. ☐

- *Let* $\mathbf{Q}$ *be an orthogonal projection on* $\mathrm{Im}(\mathbf{Q})$*, Then* $\mathbf{I}-\mathbf{Q}$*is an orthonormal projection onto* $\ker(\mathbf{Q})$.
  $\ker(\mathbf{Q})$ denotes the kernel of $\mathbf{Q}$, and $\mathrm{Im}(\mathbf{Q})$ denotes the image of $\mathbf{Q}$. First we verify the condition for a matrix to be a projection matrix:

  $$(\mathbf{I}-\mathbf{Q})^2 = (\mathbf{I}-\mathbf{Q})(\mathbf{I}-\mathbf{Q}) = \mathbf{I} - 2\mathbf{Q} + \mathbf{Q}^2 = \mathbf{I} - \mathbf{Q}.$$

  Therefore $\mathbf{I}-\mathbf{Q}$ is a projection matrix. Since $\mathbf{Q}$ is symmetric, so is $\mathbf{I}-\mathbf{Q}$ and hence an orthogonal projection. For the next par, let $\mathbf{y} \in \ker(\mathbf{Q})$, i.e., $\mathbf{Qy} = \mathbf{0}$. Then:

  $$(\mathbf{I}-\mathbf{Q})\mathbf{y} = \mathbf{y} - \mathbf{Qy} = \mathbf{y} \in \mathrm{Im}(\mathbf{I}-\mathbf{Q}).$$

  Therefore $\ker(\mathbf{Q}) \subseteq \mathrm{Im}(\mathbf{I}-\mathbf{Q})$. On the other hand, suppose that $y \in \mathrm{Im}(\mathbf{I}-\mathbf{Q})$. There is $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{y} = (\mathbf{I}-\mathbf{Q})\mathbf{x}$. We have:

  $$\mathbf{Qy} = \mathbf{Q}(\mathbf{I}-\mathbf{Q})\mathbf{x} = \mathbf{Qx} - \mathbf{Q}^2\mathbf{x} = \mathbf{Qx} - \mathbf{Qx} = \mathbf{0}.$$

  So $\mathbf{y} \in \ker(\mathbf{Q})$ and therefore $\mathrm{Im}(\mathbf{I}-\mathbf{Q}) \subseteq \ker(\mathbf{Q})$. So $\mathrm{Im}(\mathbf{I}-\mathbf{Q}) = \ker(\mathbf{Q})$.

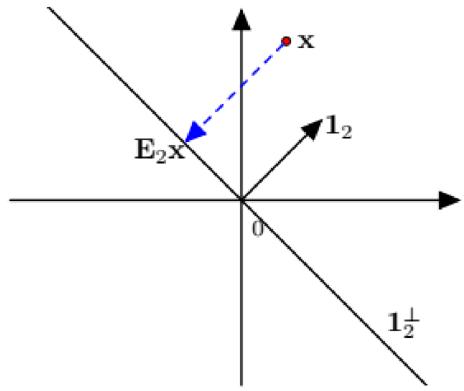

Figure 2.2: Orthogonal Projection of $\mathbf{E}_2$

- *Define* $\mathbf{E}_n$ *as follows:*

$$\mathbf{E}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_{n\times n} = \begin{bmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \dots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \dots & 1-\frac{1}{n} \end{bmatrix}$$

Then $\mathbf{E}_n$ *is an orthogonal projection onto* $\mathbf{1}_n^{\perp} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{1}_n^T\mathbf{x} = 0\}$ *where* $\mathbf{1}_n$ *is all-one-vector in* $\mathbb{R}^n$*.*
See that for all $\mathbf{x} \in \mathbb{R}^n$:

$$\mathbf{1}_n^T\mathbf{E}_n\mathbf{x} = \mathbf{1}_n^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_{n\times n})\mathbf{x} = (\mathbf{1}_n^T - \mathbf{1}_n^T)\mathbf{x} = \mathbf{0}.$$

Therefore each vector in $\mathrm{Im}(\mathbf{E}_n)$ is orthogonal to $\mathbf{1}_n$.

Note that $\frac{1}{n}\mathbf{1}_{n\times n} \times \frac{1}{n}\mathbf{1}_{n\times n} = \frac{1}{n}\mathbf{1}_{n\times n}$ and $\frac{1}{n}\mathbf{1}_{n\times n}$ is symmetric. Therefore it is an orthogonal projection. Moreover its image is a one dimensional subspace spanned by $\mathbf{1}_n$. From the previous item, $\mathbf{I}_n - \frac{1}{n}\mathbf{1}_{n\times n}$ is also an orthogonal projection onto the kernel of $\frac{1}{n}\mathbf{1}_{n\times n}$ which is $\mathbf{1}_n^{\perp}$.

Theorem 2.11 (Inverse and determinant of partitioned matrix)**.** *Let* $\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{bmatrix}$ *be a symmetric, invertible (regular) and* $\mathbf{A}$ *is also invertible (regular). Then:*

(a) The inverse matrix of $\mathbf{M}$ *is given by:*

$$\mathbf{M}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}^T & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}^T & \mathbf{E}^{-1} \end{bmatrix}$$

where $\mathbf{E}$ *is the Schur complement given by* $\mathbf{E} = \mathbf{C} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}$ *and* $\mathbf{F} = \mathbf{A}^{-1}\mathbf{B}$.

(b) The determinant of $\mathbf{M}$ *is given by:*

$$\det(\mathbf{M}) = \det(\mathbf{A})\det(\mathbf{C} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}).$$

There is also an extension of this theorem for general case where $\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$ (see [Mur12, p.118]).

Definition 2.12 (Isometry)**.** A linear transformation $\mathbf{M} : \mathbb{R}^n \to \mathbb{R}^n$ is called an isometry if $\mathbf{x}^T\mathbf{x} = (\mathbf{M}\mathbf{x})^T(\mathbf{M}\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$.

Some properties of isometries are as follows:

- If $\mathbf{U}$ and $\mathbf{V}$ are isometries, then the product $\mathbf{UV}$ is also an isometry.
- If $\mathbf{U}$ is an isometry, $|\det(\mathbf{U})| = 1$.
- If $\mathbf{U}$ is an isometry, then $|\lambda(\mathbf{U})| = 1$ for all eigenvalues of $\mathbf{U}$.