Practice 6
Orthogonality and orthogonal projection

Contents

1 Scalar product, norm and distance

We shall recall first some preliminary concepts. Let $\vec{u} = (u_1, \ldots, u_n)$ and $\vec{v} = (v_1, \ldots, v_n)$ two vectors of $\mathbb{R}^n$.

- The *scalar product* (or *dot product*) of $\vec{u}$ and $\vec{v}$ is the real number

$$\vec{u} \cdot \vec{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

  Example 1. If $\vec{u} = (1, -2, 3)$ and $\vec{v} = (-3, 4, 7)$, we can compute $\vec{u} \cdot \vec{v}$ with Scilab in the following way:

```
-->u=[1; -2; 3]; v=[-3; 4; 7];

-->u'*v
 ans  =

    10.
```

- If $\vec{u} \cdot \vec{v} = 0$ we say that $\vec{u}$ and $\vec{v}$ are *orthogonal*.

- A system of vectors $\{\vec{u}_1, \ldots, \vec{u}_r\}$ is *orthogonal* if $\vec{u}_i \cdot \vec{u}_j = 0$ whenever $i \neq j$.

- If W is a linear subspace of $\mathbb{R}^n$, the *orthogonal complement* of W is the subspace

$$W^\perp = \{z \in \mathbb{R}^n \mid z \cdot w = 0 \text{ for all } w \in W\}.$$

  If W is spanned by a set of vectors $S = \{\vec{u}_1, \ldots, \vec{u}_r\}$ then $W^\perp$ is the set of solutions of the homogeneous linear system $\mathsf{M}(S)^t\vec{x} = \vec{0}$, where $\mathsf{M}(S)$ is the matrix whose columns are the vectors of S. In other words:

$$W^\perp = \ker(\mathsf{M}(S)^t).$$

- The *column subspace* of an $m \times n$ matrix A, denoted by $Col(A)$, is the subspace of $\mathbb{R}^m$ spanned by the column vectors of A.

- The *row subspace* of an $m \times n$ matrix A, denoted by $Row(A)$, is the subspace of $\mathbb{R}^n$ spanned by the transposed row vectors of A.

  Example 2. We shall compute, with the help of Scilab, the orthogonal subspace of $W = \mathrm{span}((1,2,1,1),(2,-3,0,0))$. First we shall compute the matrix $\mathsf{M}(S)$, where $S = \{(1,2,1,1),(2,-3,0,0)\}$:

```
-->u1=[1; 2; 1; 1]; u2=[2; -3; 0; 0];

-->MS=[u1 u2]
 MS  =

    1.    2.
    2.  - 3.
    1.    0.
    1.    0.
```

  Next we compute the kernel of the matrix $\mathsf{M}(S)^t$:

```
-->kernel(MS')
 ans  =

    0.3464102    0.3464102
    0.2309401    0.2309401
  - 0.9041452    0.0958548
    0.0958548  - 0.9041452
```

  This means that

$$W^\perp = \mathrm{span}((0.3464102, 0.2309401, -0.90414, 0.0958548),$$

$$(0.3464102, 0.2309401, 0.0958548, -0.9041452))$$

- The *norm* of $\vec{v}$ is
$$\|\vec{v}\| = \sqrt{\vec{v}\cdot\vec{v}} = \sqrt{v_1^2 + \cdots + v_n^2}.$$

  Example 3. To compute the norm of a vector with Scilab we can use the command **norm**:

```
-->u=[1; 3; 8];

-->norm(u)
 ans  =

    8.6023253
```

- A vector is *unitary* if its norm is 1.
- If $\vec{v}$ is whichever non-zero vector, the vector $\frac{\vec{v}}{\|\vec{v}\|}$ is unitary.

  Example 4. We want to transform $\vec{v} = (4, -6, 9)$ into a unitary vector with the help of Scilab:

```
-->v=[4; -6; 9];

-->w=v/norm(v)
 w  =

    0.3468440
  - 0.5202660
    0.7803990

-->norm(w)
 ans  =
```

- The *distance* between $\vec{u}$ and $\vec{v}$ is $\|\vec{u} - \vec{v}\|$.

  Example 5. We compute the distance between $\vec{u} = (1, 4, -3, 1)$ and $\vec{v} = (7, -3, 2, 1)$ with Scilab:

```
-->u=[1; 4; -3; 1]; v=[7; -3; 2; 1];

-->norm(u-v)
 ans  =

    10.488088
```

- Observe that, given a real matrix A, the *kernel* of A is equal to the *orthogonal complement of the row subspace*, that is:
$$Row(\mathsf{A})^{\perp} = \ker(\mathsf{A}).$$
Since the *column subspace* of A is equal to the row subspace of A^t, one has that $Col(\mathsf{A})^{\perp} = Row(\mathsf{A}^t)^{\perp} = \ker(A^t)$. Therefore the *orthogonal complement of the column subspace* is equal to the *kernel* of A^t:
$$Col(\mathsf{A})^{\perp} = \ker(A^t).$$

Example 6. Let $\mathsf{A} = \begin{bmatrix} 5 & -4 & 3 \\ -4 & 1 & 3 \\ 0 & -3 & 5 \\ 3 & 8 & -1 \end{bmatrix}$. We shall compute, with Scilab, the subspaces $Row(\mathsf{A})^{\perp}$ and $Col(\mathsf{A})^{\perp}$:

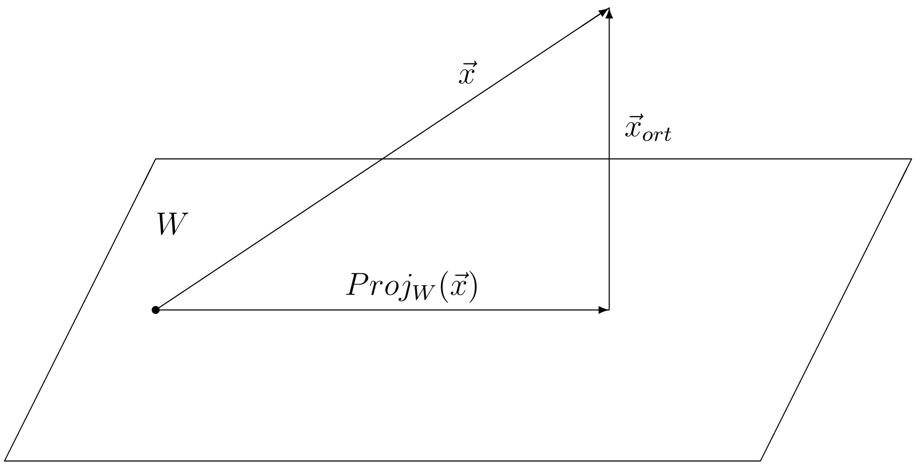
```
-->A=[5 -4 3; -4 1 3; 0 -3 5; 3 8 -1];

-->RowOrt=kernel(A)
 RowOrt  =

     []

-->ColOrt=kernel(A')
 ColOrt  =

    0.4958960
    0.5689754
  - 0.6524947
  - 0.0678594
```

This means that $Row(\mathsf{A})^{\perp} = \{\vec{0}\}$ and
$$Col(\mathsf{A})^{\perp} = \mathrm{span}((0.4958960, 0.5689754, -0.6524947, -0.0678594)).$$

2 Orthogonal projection

2.1 Definition

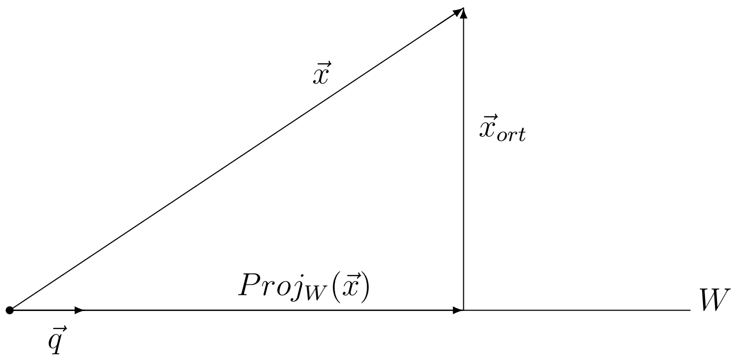
Let W be a linear subspace of $\mathbb{R}^n$. For each vector $\vec{x} \in \mathbb{R}^n$ there exists one unique vector $\vec{w} \in W$ and a unique vector of the orthogonal complement $\vec{x}_{ort} \in W^{\perp}$ such that $\vec{x} = \vec{w} + \vec{x}_{ort}$. The vector $\vec{w}$ is called *orthogonal projection* of $\vec{x}$ over the subspace W, and it will be denoted by $Proj_W(\vec{x})$. Though abstract, this definition of "orthogonal projection" formalizes and generalizes the graphical idea of "perpendicular projection":



The vector $\vec{x}$ can be decomposed, in a unique way, as a sum of two orthogonal *components*: one of them ($Proj_W(\vec{x})$) *over* W and the other ($\vec{x}_{ort}$) orthogonal to all vectors in W (that is, belonging to $W^{\perp}$). When we talk about *orthogonal projection* of $\vec{x}$ over W we are talking about the component $Proj_W(\vec{x})$ over W in that decomposition.

2.2 Easy case: orthogonal projection over a line

Let us consider the case in which W is a line in $\mathbb{R}^n$, that is, it is generated by a unique non-zero vector. We can divide this generator by its norm and transform it into a unitary vector. Therefore we can take $S = \{\vec{q}\}$, where $\vec{q}$ is unitary.



Take a vector $\vec{x} \in \mathbb{R}^n$. We want to compute the orthogonal projection $Proj_W(\vec{x})$ of $\vec{x}$ over W (see the above picture). Since $Proj_W(\vec{x})$ belongs to the line W, there exists a scalar λ such that $Proj_W(\vec{x}) = \lambda\vec{q}$. But the vector $\vec{x} - \lambda\vec{q} = \vec{x}_{ort}$ is orthogonal to W and, therefore, $(\vec{x} - \lambda\vec{q}) \cdot \vec{q} = 0$. Now $\vec{x} \cdot \vec{q} - \lambda\vec{q} \cdot \vec{q} = 0$ and, since $\vec{q}$ is unitary, one has $\lambda = \vec{q} \cdot \vec{x}$ or, using the notation of product of a row vector by a column vector instead of the scalar product notation, $\lambda = \vec{q}^{\,t}\vec{x}$. We conclude, then, the following assertion:

The projection of a vector $\vec{x}$ over a line W is the vector

$$Proj_W(\vec{x}) = (\vec{q}^{\,t}\vec{x})\vec{q}$$

where $\vec{q}$ is a unitary generator of the line.

Example 7. Let us consider the line $W = \mathrm{span}(1, -2, 5) \subseteq \mathbb{R}^3$ and we shall compute the projection of the vector $\vec{x} = (0, 1, 1)$ over W. First we compute a unitary generator of the line dividing by the norm:

```
-->u=[1; -2; 5];

-->q=u/norm(u)
 q  =

    0.1825742
  - 0.3651484
    0.9128709
```

Now we compute the projection using the above formula:

```
-->x=[0; 1; 1];

-->(q'*x)*q
 ans  =

    0.1
  - 0.2
    0.5
```

Then, the orthogonal projection is $(0.1, -0.2, 0.5)$.

2.3 General case

Now we are going to describe how to compute the orthogonal projection of a vector over a subspace W. Assume that we know a system of generators $S = \{\vec{u}_1, \ldots, \vec{u}_r\}$ of W. Then W can be seen as the column subspace of the matrix $\mathsf{M}(S)$ (the matrix whose columns are the vectors $\vec{u}_i$). Therefore

$$W^\perp = \ker(\mathsf{M}(S)^t).$$

Since $\vec{x} = Proj_W(\vec{x}) + \vec{x}_{ort}$ one has that $\vec{x} - Proj_W(\vec{x}) = \vec{x}_{ort}$ and, therefore,

$$\boxed{\vec{x} - Proj_W(\vec{x}) \text{ must be orthogonal to } W,}$$

that is,

$$\vec{x} - Proj_W(\vec{x}) \in W^{\perp} = \ker(\mathsf{M}(S)^t).$$

Hence

$$\mathsf{M}(S)^t(\vec{x} - Proj_W(\vec{x})) = \vec{0},$$

that is,

$$\mathsf{M}(S)^t Proj_W(\vec{x}) = \mathsf{M}(S)^t\vec{x}. \tag{1}$$

On the other hand

$$\boxed{\text{the vector } Proj_W(\vec{x}) \text{ belongs to } W}$$

and this implies that it can be written as a linear combination of the vectors of S (which is a system of generators of W). Denote by

$$Proj_W(\vec{x}) = y_1\vec{u}_1 + \cdots + y_r\vec{u}_r \tag{2}$$

a linear combination of this type and by

$$\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_r \end{bmatrix}$$

the vector of coefficients. Equality (2) means that

$$Proj_W(\vec{x}) = \mathsf{M}(S)\vec{y}.$$

Replacing this new expression of $Proj_W(\vec{x})$ in Equality (1) one has

$$M(S)^t\mathsf{M}(S)\vec{y} = \mathsf{M}(S)^t\vec{x}.$$

Notice that the matrix $M(S)^t\mathsf{M}(S)$ is a square matrix of order r. This is the matrix expression of a system of r linear equations with r unknowns such that $\vec{y}$ is the vector of unknowns and $M(S)^t\vec{x}$ is the vector of independent terms (remember that we know $\vec{x}$!). It can be reasoned that all the solutions $(y_1, \ldots, y_r)$ of this system give rise to the same linear combination $y_1\vec{u}_1 + \cdots + y_r\vec{u}_r$, which is the desired projection.

We have deduced the following result:

> The projection of a vector $\vec{x}$ over the linear subspace W generated by a set $S = \{\vec{u}_1, \ldots, \vec{u}_r\}$ is the vector $Proj_W(\vec{x}) = y_1\vec{u}_1 + \cdots + y_r\vec{u}_r$, where $\vec{y} = (y_1, \ldots, y_r)$ is a solution of the system
>
> $$M(S)^t\mathsf{M}(S)\vec{y} = \mathsf{M}(S)^t\vec{x}. \tag{3}$$

Assume now that S is linearly independent, that is, it is a basis of W.

In this case, $rank(\mathsf{M}(S)^t\mathsf{M}(S)) = r$ and this implies that the system (3) **has a unique solution**, which gives the coefficients $\vec{y}_i$ of the orthogonal projection of $\vec{x}$ with respect to the basis S. Since the matrix $\mathsf{M}(S)^t\mathsf{M}(S)$ is invertible we can compute easily the solution:

$$\vec{y} = (\mathsf{M}(S)^t\mathsf{M}(S))^{-1}\mathsf{M}(S)^t\vec{x}.$$

Since the components of $\vec{y}$ are the coefficients of the projection with respect to the basis S, the product $\mathsf{M}(S)\vec{y}$ will be equal to the desired projection $Proj_W(\vec{x})$:

$$Proj_W(\vec{x}) = \mathsf{M}(S)(\mathsf{M}(S)^t\mathsf{M}(S))^{-1}\mathsf{M}(S)^t\vec{x}.$$

So, we have obtained a very nice formula to compute orthogonal projections!

The projection of a vector $\vec{x}$ over the linear subspace W generated by a **linearly independent** set S is

$$Proj_W(\vec{x}) = \mathsf{P}_W\vec{x}, \tag{4}$$

where

$$\mathsf{P}_W = \mathsf{M}(S)(\mathsf{M}(S)^t\mathsf{M}(S))^{-1}\mathsf{M}(S)^t$$

will be called the *projection matrix* over W.

It can be seen that, if a different basis S is taken, the projection matrix does not changes. Then the notation P_W is correct, since P_W depends only on W, and not on the particular choice of the basis S.

Example 8. Let W be the linear subspace of $\mathbb{R}^3$ with basis $S = \{(1,2,3), (-3,5,1)\}$ and consider the vector $\vec{x} = (2,3,4) \in \mathbb{R}^3$. We shall compute, with the help of Scilab, the orthogonal projection of $\vec{x}$ over W.

First, we define the matrix $\mathsf{M}(S)$:

```
-->u1=[1; 2; 3;]; u2=[-3; -5; 1]; MS=[u1 u2]
 MS  =

    1.  - 3.
    2.  - 5.
    3.    1.
```

We compute the projection matrix P_W:

```
-->x=[2; 3; 4];
-->PW=MS*inv(MS'*MS)*MS'
 PW  =

    0.2589744    0.4358974  - 0.0435897
    0.4358974    0.7435897    0.0256410
  - 0.0435897    0.0256410    0.9974359
```

Multiplying this matrix by the vector $\vec{x}$ we obtain the orthogonal projection:

```
-->PW*x
 ans  =

    1.6512821
    3.2051282
    3.9794872
```

Therefore $Proj_W(\vec{x}) = (1.6512821, 3.2051282, 3.9794872)$.