Applications of Orthogonal Projections

Niklas Walter*

1 Recap on Projections

We start by recalling the definition of (orthogonal) projections on a general vector space V or Hilbert space $\mathcal{H}$ respectively.

Definition 1.1. Let V be a vector space. A projection P is an idempotent endomorphism on the V, i.e

$$P : V \to V \text{ and } P^2 = P.$$

A particularly interesting case of projections, are orthogonal projections. To get idea about orthogonality we need to introduce an inner product. Hence, we consider a Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ equipped with an inner product. In particular, we call two vectors $u, v \in \mathcal{H}$ orthogonal if $\langle u, v \rangle = 0$. Recall that the norm on $\mathcal{H}$ is induced by $\langle \cdot, \cdot \rangle$ in the following way

$$\|u\| := \sqrt{\langle u, u \rangle},\ u \in \mathcal{H}.$$

Based on the inner product, we can now introduce a notion for a projection being orthogonal. In particular, a projection $P : \mathcal{H} \to \mathcal{H}$ is called orthogonal if it satisfies for all $u, v \in \mathcal{H}$

$$\langle Pu, v \rangle = \langle u, Pv \rangle.$$

To characterise a orthogonal projection, we define for a subspace $\mathcal{G} \subset \mathcal{H}$ its orthogonal complement by

$$\mathcal{G}^\perp := \{u \in \mathcal{H} \,:\, \langle u, v \rangle \text{ for all } v \in \mathcal{G}\}.$$

Theorem 1.2 (Projection Theorem)**.** Let $\mathcal{G} \subset \mathcal{H}$ be a Hilbert subspace and $u \in \mathcal{H}$. Then

1. there exists a unique $\hat{u} \in \mathcal{G}$ such that

$$\|u - \hat{u}\| = \inf_{v \in \mathcal{G}} \|u - v\|,$$

2. the vector $\hat{u}$ is uniquely characterised by $u - \hat{u} \in \mathcal{G}^\perp$.

2 Ordinary Least Squares (OLS)

Consider now the Hilbert space $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ for some $n \in \mathbb{N}$ with the Euclidean inner product defined by

$$\langle x, y \rangle := \langle x, y \rangle_{\text{eucl.}} = x^T y = \sum_{i=1}^{n} x_i y_i,$$

*The author assumes no liability for the correctness of the following content. For comments and inspiration please contact: walter@math.lmu.de

where x_i and y_i denote the i-th entry of x and y respectively. Therefore, the induced Euclidean norm is given by

$$\|x\| := \|x\|_{\text{eucl.}} = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Then for some $y \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n\times p}$ the ordinary least square problem wants to solve

$$\min_{\beta\in\mathbb{R}^p} \|y - X\beta\|^2.$$

Hence, we want to find the vector $\hat{y}$ in the column space X minimising the distance between y and the subspace spanned by the columns of X. In other words, we want to project the vector y onto the subspace

$$\mathfrak{M}_X := \left\{ X\beta = \sum_{i=1}^{p} \beta_i X_i,\ \beta \in \mathbb{R}^p \right\} \subseteq \mathbb{R}^n.$$

Note that $\mathfrak{M}_X$ is a Hilbert subspace. Hence, by the projection theorem we get that the error vector

$$\varepsilon := y - \hat{y} = y - X\hat{\beta}$$

must be orthogonal to every vector in $\mathfrak{M}_X$. In particular, ε is orthogonal to every column of X. Formally, we thus obtain

$$X^T\varepsilon = X^T(y - X\hat{\beta}) \overset{!}{=} 0.$$

Assuming X^TX to be invertible we get

$$\hat{y} = X\hat{\beta} = X(X^TX)^{-1}X^Ty =: P_X y.$$

Note that

$$P_X^2 = X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = P_X$$

and for every $v, w \in \mathbb{R}^n$ we observe

$$\langle P_X v, w\rangle = (P_X v)^T w = v^T P_X^T w = v^T P_X w = \langle v, P_X w\rangle.$$

Therefore, the matrix P_X is indeed the orthogonal projection onto the column space of X. This demonstrates that heuristically the OLS/linear regression is nothing else than projection the response vector onto the space spanned by the regressor variables.

3 Conditional expectation

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then we define the vector space of square-integrable random variables as

$$\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P}) := \left\{ X : \Omega \to \mathbb{R} \,\middle|\, X \in \mathcal{F},\ \mathbb{E}[X^2] = \int_\Omega X^2(\omega)\,\mathrm{d}\mathbb{P}(\omega) < \infty \right\}.$$

Moreover, we introduce the seminorm $\|\cdot\|_{\mathcal{L}^2} : \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ as

$$\|X\|_{\mathcal{L}^2} := \mathbb{E}[X^2]^{1/2} = \left(\int_\Omega X^2(\omega)\,\mathrm{d}\mathbb{P}(\omega)\right)^{1/2}. \tag{3.1}$$

The seminorm $\|\cdot\|_{\mathcal{L}^2}$ becomes a norm if and only if the empty set $\emptyset$ is the only null-set in $\mathcal{F}$. This is easy to see, because if the is a null-set $\emptyset \neq N \in \mathcal{F}$ it would follow $\|\mathbb{1}_N\|_{\mathcal{L}^2} = 0$ without $\mathbb{1}_N$ being th null function. Therefore, we introduce the equivalence relation

$$X \sim Y :\Longleftrightarrow \mathbb{P}(X = Y) = 1.$$

So we identify two random variables to be equivalent if they are equal $\mathbb{P}$-almost surely. Based on this we can define the space

$$L^2(\Omega, \mathcal{F}, \mathbb{P}) := \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})/\sim,$$

which is now a Hilbert space with an inner product defined as

$$\langle X, Y\rangle_{L^2} := \mathbb{E}[XY]^{1/2} = \left(\int_\Omega X(\omega)Y(\omega)\,\mathrm{d}\mathbb{P}(\omega)\right)^{1/2}.$$

We can directly see that the regarding norm $\|\cdot\|_{L^2}$ is computed in the same way as the seminorm (3.1) on $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$.

We now focus on how the conditional expectation can be deduced as the result of an orthogonal projection of a random variables in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ onto a suitable subspace. In particular, we fix two random variable $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. For the Y we now want to solve the optimisation problem

$$\min_{L\in\mathfrak{L}} \|Y - L\|^2_{L^2}$$

for the subspace

$$\mathfrak{L} := \left\{L \in L^2(\Omega, \mathcal{F}, \mathbb{P}) : \ L \text{ is } \sigma(X)\text{-measurable}\right\}.$$

By the Doob-Dynkin Lemma, the measurability implies that every $L \in \mathfrak{L}$ can be written as $L = f(X)$ for a measurable deterministic real-valued function f. Then by the projection theorem there exists a $\hat{Y} \in \mathfrak{L}$ such that

$$\|Y - \hat{Y}\|^2_{L^2} = \min_{L\in\mathfrak{L}} \|Y - L\|^2_{L^2},$$

which is (uniquely) characterised by the identity

$$\langle Y - \hat{Y}, L\rangle^2_{L^2} = 0,$$

which must hold for all $L \in \mathfrak{L}$. In other words, for every $\sigma(X)$-measurable random variable $L \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ we obtain

$$\mathbb{E}[YL] = \mathbb{E}[\hat{Y}L].$$

By its definition we know that this equality is solved by the conditional expectation

$$\hat{Y} = \mathbb{E}[Y|X].$$

Hence, in the space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ the conditional expectation is nothing else but the orthogonal projection of Y onto the subspace of all $\sigma(X)$-measurable random variables. For this reason, the conditional expectation in this setting is also called regression.

4 Fourier Analysis

In this section, we derive a nice connection between orthogonal projections and the concept of Fourier series of periodic functions. In Section 4.1, we consider only real-valued periodic functions and arrive at the so called *since-cosine form*. In Section 4.2, we extend the class of functions to complex valued functions with some assumptions on the integrability property to derive the *exponential form*.

4.1 Fourier series for $\mathbb{R}$-valued functions

Here, we consider the space $\mathcal{C}_{2\pi}([-\pi,\pi],\mathbb{R})$ of continuous and 2π-periodic functions, i.e.

$$\mathcal{C}_{2\pi}([-\pi,\pi],\mathbb{R}) := \{f : [-\pi,\pi] \to \mathbb{R} \mid f \text{ continuous and } f(x) = f(x+2\pi)\}\,.$$

This space is a Hilbert space together with the inner product defined as

$$\langle f, g\rangle := \langle f, g\rangle_{\mathcal{C}_{2\pi}([-\pi,\pi],\mathbb{R})} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\,\mathrm{d}x$$

for any $f, g \in \mathcal{C}([-\pi,\pi],\mathbb{R})$. We denote by $\|\cdot\| := \|\cdot\|_{\mathcal{C}_{2\pi}([-\pi,\pi],\mathbb{R})}$ the corresponding induced norm. Further, for some $n \in \mathbb{N}\cup\{+\infty\}$ we consider the subspace $B \subset \mathcal{C}_{2\pi}([-\pi,\pi],\mathbb{R})$ defined by

$$B := \{1, \cos(kx), \sin(kx) \mid k = 1, ..., n\}\,.$$

For any $m, l \in \{1, ..., n\}$ it holds by the periodic property

$$\langle \sin(mx), \cos(lx)\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \sin(mx)\cos(lx)\,\mathrm{d}x = \begin{cases} 0, & \text{if } m \neq l,\\ 1, & \text{if } m = l,\end{cases}$$

and

$$\langle \sin(mx), \sin(lx)\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \sin(mx)\sin(lx)\,\mathrm{d}x = \begin{cases} 0, & \text{if } m \neq l,\\ 1, & \text{if } m = l,\end{cases}$$

and

$$\langle \cos(mx), \cos(lx)\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos(mx)\cos(lx)\,\mathrm{d}x = \begin{cases} 0, & \text{if } m \neq l,\\ 1, & \text{if } m = l.\end{cases}$$

So that the functions contained in B form a orthonormal system. Our aim now is to project a 2π-periodic function $f \in \mathcal{C}_{2\pi}([-\pi,\pi],\mathbb{R})$ onto the subspace spanned by the function in B given by

$$\operatorname{span}(B) := \left\{ h \in \mathcal{C}_{2\pi}([-\pi,\pi]) \;\middle|\; h(x) = a_0 + \sum_{k=1}^{n}[a_k\sin(kx) + b_k\cos(kx)] \right\}.$$

By the projection theorem there exists a $\hat{f} \in \operatorname{span}(B)$ solving the opimisation problem in the sense that

$$\|f - \hat{f}\| = \min_{h\in\operatorname{span}(B)} \|f - h\|,$$

which is uniquely characterised by the identity

$$\langle f - \hat{f}, h\rangle = 0 \text{ for all } h \in \operatorname{span}(B).$$

In particular, we obtain the following system of equations which we can use to derive the coefficients of the approximating function $\hat{f}$

$$\begin{cases} \langle f - \hat{f}, 1 \rangle = 0, \\ \langle f - \hat{f}, \cos(kx) \rangle, \ k = 1, ..., n, \\ \langle f - \hat{f}, \sin(kx) \rangle, \ k = 1, ..., n. \end{cases}$$

From the first equation we obtain

$$\begin{aligned} \langle f - \hat{f}, 1 \rangle &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \, \mathrm{d}x - \left(\langle \hat{a}_0, 1 \rangle + \sum_{k=1}^{n} [\hat{a}_k \langle \sin(kx), 1 \rangle + \hat{b}_k \langle \cos(kx), 1 \rangle] \right) \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \, \mathrm{d}x - 2\hat{a}_0 \\ &\stackrel{!}{=} 0. \end{aligned}$$

Therefore, we finally get

$$\hat{a}_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \, \mathrm{d}x.$$

In the same manner, using the orthonormality we arrive at

$$\hat{a}_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(kx) \, \mathrm{d}x \text{ and } \hat{b}_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(kx) \, \mathrm{d}x.$$

The function $\hat{f}$ is the Fourier series in the cosine-sine form of the function f given by

$$\hat{f}(x) = \hat{a}_0 + \sum_{k=0}^{n} [\hat{a}_k \sin(kx) + \hat{b}_k \cos(kx)].$$

4.2 Fourier series for $\mathbb{C}$-valued functions

In this section, we follow a similar approach to derive the Fourier based based on the ideas of orthogonal projections. This time we allow the approximated function to attain values in $\mathbb{C}$. As before, we consider functions with period 2π. The following results can easily be extended to a general period $L > 0$. We can derive a nice geometric/topological interpretation of 2π-periodic functions. We denote by $\mathbb{S}^1$ the unit circle. Then we note that the function $\iota : \mathbb{R} \to \mathbb{S}^1$ defined as $\iota : x \mapsto e^{ix}$ is surjective morphism. Moreover, we get by Euler's formula that it has the kernel $2\pi\mathbb{Z}$. Hence, by considering the factor space $\mathbb{R}/2\pi\mathbb{Z}$ as domain we can turn ι into a invective and thus bijective function. This observation can be summarised in the following commutative diagram

$$\overset{\hat{\iota}}{\cong}$$

Therefore, every 2π-periodic function can be characterised by a function $f : \mathbb{S}^1 \to \mathbb{C}$. In the following, we introduce the space

$$\mathcal{L}^2(\mathbb{S}^1) := \left\{ f : \mathbb{S}^1 \to \mathbb{C} \;\middle|\; f \text{ is Lebesgue-integrable and } \int_{\mathbb{S}^1} |f(x)|^2 \, \mathrm{d}x < \infty \right\}.$$

Like in Section 3 we obtain the vector space $L^2(\mathbb{S}^1)$ by defining $L^2(\mathbb{S}^1) := \mathcal{L}^2(\mathbb{S}^1)/\sim$ with the equivalence relation defined as

$$f \sim g :\Longleftrightarrow f = g \text{ almost everywhere.}$$

The space $L^2(\mathbb{S}^1)$ is turned into a Hilbert space together with the inner product

$$\langle f, g \rangle := \langle f, g \rangle_{L^2(\mathbb{S}^1)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{g(x)}\, \mathrm{d}x$$

with $f, g \in L^2(\mathbb{S}^1)$ and with $\bar{g}$ denoting the complex conjugate of the function g. We denote by $\|\cdot\| := \|\cdot\|_{L^2(\mathbb{S}^1)}$ the corresponding induced norm. For a fixed $n \in \mathbb{N}_0 \cup \{+\infty\}$ we now consider the subset $B' \subset L^2(\mathbb{S}^1)$ defined by

$$B' := \left\{ e^{i2\pi kx} \,\middle|\, -n \leq k \leq n \right\}.$$

Note that the functions contained in B' are orthonormal since for any $m, l \in \mathbb{N}_0 \cup \{+\infty\}$ it holds by the periodic property

$$\left\langle e^{i2\pi mx}, e^{i2\pi lx} \right\rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i2\pi(m-l)x}\, \mathrm{d}x = \begin{cases} 0, & \text{if } m \neq l, \\ 1, & \text{if } m = l. \end{cases}$$

Analog to the previous section, we want to project a function $f \in L^2(\mathbb{S}^1)$ to the space spanned by the functions in B' given by

$$\operatorname{span}(B') = \left\{ h \in L^2(\mathbb{S}^1) \,\middle|\, h(x) = \sum_{k=-n}^{n} c_k e^{i2\pi kx} \right\}.$$

Again by the projection theorem there exists a $\hat{f} \in \operatorname{span}(B')$ solving the optimisation problem in the sense that

$$\|f - \hat{f}\| = \min_{h \in \operatorname{span}(B')} \|f - h\|,$$

which is uniquely characterised by the identity $\langle f - \hat{f}, h \rangle = 0$ for all $h \in \operatorname{span}(B')$. In particular, it must hold for every $k = -n, ..., n$ that

$$\left\langle f - \hat{f}, e^{i2\pi kx} \right\rangle = 0.$$

In particular, we obtain for the coefficients $\hat{c}_k$ of the function $\hat{f}$

$$\begin{aligned} & \left\langle f, e^{i2\pi kx} \right\rangle = \left\langle \hat{f}, e^{i2\pi kx} \right\rangle \\ \Longleftrightarrow\ & \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{i2\pi kx}\, \mathrm{d}x = \sum_{l=-n}^{n} \hat{c}_l \left\langle e^{i2\pi lx}, e^{i2\pi kx} \right\rangle \\ \Longleftrightarrow\ & \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{i2\pi kx}\, \mathrm{d}x = \sum_{l=-n}^{n} \hat{c}_l \mathbb{1}_{k=l} \\ \Longleftrightarrow\ & \hat{c}_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{i2\pi kx}\, \mathrm{d}x. \end{aligned}$$

Hence, we arrive at the Fourier series in the (complex) exponential form of the function f given by

$$\hat{f}(x) = \sum_{k=-n}^{n} \hat{c}_k e^{-i2\pi kx}.$$