Projection and approximation

Yinan Huang

Contents

1 Preface **1**

2 Projection operator **1**
2.1 Projection in an intuitive way 1
2.2 Projection in a formal way 2
2.3 Orthogonal projection and projection theorem 3

3 Projection approximation in Hilbert space **4**
3.1 Normal equation 4
3.2 Approximation problems in representation 5
3.3 Examples of approximation 5
3.3.1 Polynomials approximation 5
3.3.2 Linear regression 6
3.3.3 Least square filtering 6

4 Dual approximation problem **7**

5 Conclusion **7**

1 Preface

Projection is a very striking and useful concept in engineering application. Plenty of approximation and estimation (optimization) problems is acutually doing projection to minimize “error”. For example, deciding coefficients of polynomials for approximation of signals, estimating of parameters in linear regression, these are all about projection in a vector space point of view. This note will begin with introducing the definition and properties of projection operator (We assume readers have fundamental knowledge of vector space.), and gives some applications in approximation and estimation, including polynomials approximation, linear regression, etc. Finally we mention the dual problem of projection approximation.

2 Projection operator

2.1 Projection in an intuitive way

It will be helpful and inspiring to first look at projection in an intuitive way, though projection has a rigorous mathematical definition that will be introduced later. When we talk about projection, we project a vector x in vector space S to a low-dimentional subspace V. For example, in $\mathbb{R}^2$ space, a vector x can be projected to a line V (subspace $\mathbb{R}$). And also be awared that the projection is not simply determined by projection space V, but also another disjoint space W. Here two disjoint space means $V \cap W = \{0\}$. In the case of fixing projection space V, we still have many choices to do the projection by choosing different W. We generalize this to a naive

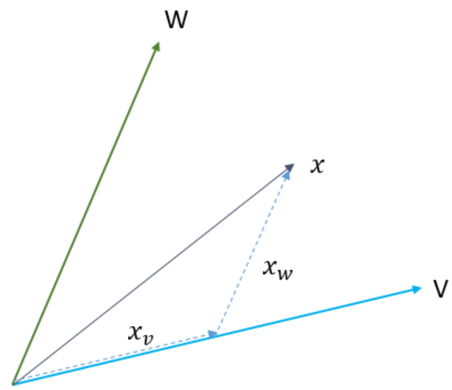

Figure 1: Projection in $\mathbb{R}^2$

definition of projection as follows,

Definition 2.1. *Let S be a vector space. A linear operator P: $S \to S$ defined on two disjoint subspaces V, W with $S = V + W$ is called a projection, if for any vector $x \in S$, it reads out the V component of x:*

$$x = v + w, \quad v \in V, w \in W \tag{1}$$

$$P(x) = v. \tag{2}$$

Note that this definition works when v and w are unique for any given x (or the projection can lead to different results). This is equivalent to say subspace V and W are linear independent. Acually we can see that it is true by the following theorem.

Theorem 2.1. *If two subspace V and W are disjoint and $S = V + W$, then for any $x \in S$, x can be uniquely decomposed into $x = v + w$ with $v \in V$ and $w \in W$. We also call V and W such that $V \cap W = \{0\}$ and $S = V + W$ are algebraic complement in S to each other.*

Proof 2.1. *Suppose $S = Span\{e_1, e_2, ..., e_n\} = \{x \in S | x = \sum_{i=0}^{n} x_i e_i\}$. By requiring subspace V and W to be algebraic complement, we must assign basis e_i to V and W without overlapping (obvious via contradiction):*

$$V = \{v \in S | v = \sum_{i=0}^{k} v_i e_i\}, \quad W = \{w \in S | w = \sum_{i=k+1}^{n} w_i e_i\}. \tag{3}$$

Thus by the linear independence of basis, we have a unique representation of any $x \in S$

$$x = \sum_{i=0}^{n} x_i e_i = \sum_{i=0}^{k} x_i e_i + \sum_{i=k+1}^{n} x_i e_i \equiv v + w. \tag{4}$$

In conclusion, we naively view projection P as: (1) Decomposes vector space S into a projection space V and its algebraic complement W; (2) Acts on a vector x, reads out the component v in V. But how to determine V and W? We are going to give a formal way of defining projection in a second.

2.2 Projection in a formal way

The naive definition of projection requires we define two subspaces V and W. But we actually have a formal and concise way to do the same thing. Note that if we project twice, we will have the same result with simply doing projection once. In a mathematical language, we have

$$P^n = P, \quad n = 1, 2, 3, ... \tag{5}$$

So fomally we can define a general projection by this idempotent property.

Definition 2.2. *A linear operator $P : S \to S$ is called a projection, if $P^2 = P$.*

This is an abstract (thus powerful) definition, at the first glance. Where is the projection space V and its algebraic complement W? To find V and W, let us first define range space and null space.

Definition 2.3. *For a operator $L : S \to S$, its range space is define by $R(L) \equiv \{y \in S | \exists x \in S s.t. L(x) = y\}$. Its null space is defined by $N(L) \equiv \{x \in S | L(x) = 0\}$.*

We will see (in the following theorem) that it turns out that for projection operator, its range space $R(P)$ is exactly the projection space V we mentioned, and its null space is just the algebraic complement of V. The range space $R(P)$ along with the null space $N(P)$ uniquely define a projection P.

Theorem 2.2. *For projection $P : S \to S$, any vector $x \in S$ can be uniquely decomposed into $x = v + w$, where $v \in R(P)$ and $w \in N(P)$. For any $x \in S$, we then have $P \cdot x = v$.*

Proof 2.2. *Note that for any $x \in S$, we have the equality $x = P \cdot x + (1 - P) \cdot x$. According to the definition of range space, it is true that $P \cdot x \in R(P)$. Remember the definition of projection $P^2 = P$, then we have $P(1 - P) \cdot x = (P - P) \cdot x = 0$, which implies $(1 - P) \cdot x \in N(P)$. Besides, we need to prove the uniqueness by showing $R(P)$ and $N(P)$ is algebraic complement. Statement $S = R(P) + N(P)$ is obvious by the fact $x = P \cdot x + (1 - P) \cdot x$. And suppose $x \in R(P) \cap N(P)$, then $\exists y \in S s.t. P \cdot y = x$ and $P \cdot x = 0$. Apply P by both sides, we have $0 = P^2 \cdot y = P \cdot y = x$, which indicated that $x = 0$. Thus $R(P) \cap N(P) = \{0\}$. In a word, $R(P)$ and $N(P)$ is algebraic complement to each other. Thus based on Thereom 1.1, we prove the uniqueness.*

In a word, the range space $R(P)$ and null space $N(P)$ are exactly the V and W we mentioned in the intuitive way of understanding projection. In vector space S, given an algebraic complement pairs V and W, we can always reads out component v of $x = v + w$ and kill component w. And what a projection operator P does is to specify V and W, that is $V = R(P)$ and $W = N(P)$. Any decomposition $S = V + W$ with $V \cap W = \{0\}$ stands for certain projection P.

2.3 Orthogonal projection and projection theorem

In practice, we are more interested in a special type of projection, called orthogonal projection.

Definition 2.4. *A projection is called an orthogonal projection, if $R(P) \perp N(p)$.*

Here by definition, two orthogonal space $V \perp W$ is a binary relation such that for any $v \in V$, $w \in W$, we have $\langle v, w \rangle = 0$. A orthogonal prjection implies that its range space $R(P)$ and null space $N(P)$ keep a special relation: they are orthogonal complement to each other. The reason why orthogonal projection is important is its role in projection theorem, which we will cover in a second.

Theorem 2.3. ***Projection Theorem*** *Let S be a Hilbert space and V be a closed subspace of S. For any vector $x \in S$, there exists a unique $v_0 \in V$, such that $\forall v \in V$, $\|x - v_0\| \leq \|x - v\|$. Furthermore, we have $x - v_0 \perp v$ for all $v \in V$.*

The projection theorem indicates that when we choose a orthogonal projection, the "error" vector $w = x - v$ will be minimized in terms of its norm. Let's prove this result.

Proof 2.3. *We are not going to prove the existence of v_0. Let's assume v_0 exists, and prove it is equivalent to the orthogonality of projection. Let us first suppose v_0 mimimize $\|x - v\|$. Let $\langle x - v_0, v\rangle = \delta$ for a vector $v \in V$, then consider vector $z = v_0 + \delta \cdot v$, then*

$$\|x - z\|^2 = \langle x - v_0 - \delta \cdot v, x - v_0 - \delta \cdot v\rangle = \|x - v_0\|^2 - 2|\delta|^2 + |\delta|^2 \leq \|x - v_0\|^2 . \quad (6)$$

Since $v = v_0$ minimizes $\|x - v\|$, so we must have $\delta = 0$. Thus $x - v_0 \perp V$. Conversely, let us suppose $x - v_0 \perp V$, then for any vector $v \in V$, it reads

$$\|x - v\|^2 = \|x - v_0 + v_0 - v\|^2 \leq \|x - v_0\|^2 + \|v_0 - v\|^2 \geq \|x - v_0\|^2 , \quad (7)$$

where in the second step from the end we use orthogonality of $x - v_0$.

In conclusion, given a projection space $V = R(P)$, there are many choices to select complement space $W = N(P)$ to meet $S = V + W$ (and thus determine a specific P). One important choose of them is orthogonal projection, that is $S = V + V^{\perp}$, which minimizes $\|x - P \cdot x\|$. Orthogonal projection is important because it minmizes the norm of the "error" $e = x - P \cdot x$, which is the goal of many practical problems.

3 Projection approximation in Hilbert space

3.1 Normal equation

An approximation problem in Hilbert space is described in the following: in Hilbert space S, given a set of vectors $\{p_i\}$ (called data vectors) and a vector $x \in S$, how to use $\hat{x} = \sum_i c_i p_i$ to approximate x s.t. error length $\|e\|^2 = \|x - \sum_i c_i p_i\|^2$ is minimized? Obviously, via projection theorem, we conclude that error e must be perpendicular to the data space (projection space) spanned by $\{p_i\}$, which means for any p_j,

$$\langle e, p_j\rangle = \langle x - \sum_i c_i p_i, p_j\rangle = 0. \quad (8)$$

Thus we obtain the Normal equation

$$\sum_i c_i \langle p_i, p_j\rangle = \langle x, p_j\rangle. \quad (9)$$

Solving the normal equation (namely, orthogonal projection conditions) can give us $\{c_i\}$, which show what the orthogonal projection exactly is. By defining Grammian matrix $R_{i,j} \equiv \langle p_j, p_i\rangle$ and $p_{x,j} \equiv \langle x, p_j\rangle$, we can recast the normal equation to a matrix form

$$Rc = p_x. \quad (10)$$

By solving matrix equation $Rc = p_x$, we can give the approximation (estimation) of x, that is $\hat{x} = \sum_i c_i p_i$ along with the minimized error $e_{\text{min}} = x - \hat{x}$. For example, if $\{p_i\}$ is linear independent, then R is inversible, $c = R^{-1} p_x$. Thus $\hat{x} = \sum_i \left(R^{-1} p_x\right)_i p_i$.

3.2 Approximation problems in representation

In practical problem, we usually have a natural representation of the vector space, which means vectors and operators are in the form of matrix. Suppose we are given certain representation, and thus x, $\{p_i\}$ can be expressed as column vector. Then we define a matrix

$$A \equiv [p_1, p_2, ..., p_n], \tag{11}$$

where p_j here is in the column matrix representation. Then an approximation can be written in matrix form

$$x = Ac + e. \tag{12}$$

Let us clearly define an approximation problem in representation.

Definition 3.1. *We call the following problem an approximation problem: given matrix equation*

$$x = Ac + e, \tag{13}$$

where x and A are known, we are required to find column vector c s.t. $\|e\|^2 = e^\dagger We$ is minimized. W is the weight matrix depending on the problem.

We know that to minimize $\|e\|^2$, we must solve normal equation $Rc = p_x$. Let $W = I$ for simplicity. In the case we have representation, we find $R = A^\dagger A$, and $p_x = A^\dagger x$. So the normal equation in representation is

$$A^\dagger Ac = A^\dagger x. \tag{14}$$

If $\{p_i\}$ is linear independent, then

$$c = (A^\dagger A)^{-1} A^\dagger x \tag{15}$$

and the estimation $\hat{x}$ is

$$\hat{x} = Ac = A(A^\dagger A)^{-1} A^\dagger x. \tag{16}$$

It shows that in representation we can directly give projection matrix $P = A(A^\dagger A)^{-1} A^\dagger$.

3.3 Examples of approximation

There are plenties of problems that turn out to be approximation in Hilbert space. The key feature of these problems is that they all use the linear combination of some "vectors" to estimate another "vector", and the goal is to minimized the "error". Acutally all practical problems have a natural representation, but some of them are uncountable-infinite and thus not convenient to use. We will see a couple of examples soon.

3.3.1 Polynomials approximation

The problem is, given function $f(t)$ define on interval $[a, b]$, how to find a fucntion $p(t)$ constructed by linear combination of limited number (let's say m) of polynomials to approximate $f(t)$ such that error

$$\int_a^b |f(t) - p(t)|^2 dt \tag{17}$$

is minimized.

Let us transfer this problem into a rigorous expression in vector space language. Let vector space $S = L_2[a, b]$, which defines a L_2 inner product $\langle f, g\rangle = \int_a^b f(t)g^*(t)dt$ and contains all fucntions with finite norm. Then for a vector $f(t) \in S$, we need to use m vectors

$\{1, t, t^2, ..., t^{m-1}\}$ to estimate $f(t)$. The estimation is given by

$$f(t) = \sum_{i=1}^{m} c_i t^{i-1} + e(t). \tag{18}$$

We need to find c_i to minimized $\|e(t)\|^2$. Apparently this a typical form of approximation problem. span$\{1, t, t^2, ..., t^{m-1}\}$ is a projection space of $S = \text{span}\{1, t, t^2, ...\}$ with infinite number of polynomials. The normal equation reads

$$Rc = p_f, \tag{19}$$

where $R_{i,j} = \langle p_j, p_i \rangle = \int_a^b t^{i+j-2} dt = (b^{i+j-1} - a^{i+j-1})/(i + j - 1)$, and $p_{x,j} = \langle f, p_j \rangle = \int_a^b f(t) t^{j-1} dt$. Solving normal equation gives us c_i and thus we have minimum error estimation $\hat{f}(t) = g(t) = \sum_{i=1}^{m} c_i t^{i-1}$.

3.3.2 Linear regression

Suppose we have data (x_i, y_i), and we build a linear model

$$y = \alpha + \beta x + e. \tag{20}$$

α, β is the parameters we want to find, and e is some random error. Here we deal with linear regression in a style without any statistic models. Our goal is to minimize $\sum_i e_i^2$, and this kind of linear regression is called least square method. we rewrite the model as

$$Y = Ac + e, \tag{21}$$

with Y be a column vector s.t. $Y_i = y_i$, $A_{i,1} = 1$ and $A_{i,2} = x_i$. The parameter vector $c = (\alpha \quad \beta)^T$. This form is exactly the approximation problem in representation. We can conclude the linear regression problem in the vector space language. In Hibert space $(\mathbb{R}^n, L_2)$ with L_2 inner product $\langle x, y \rangle = \sum_i x_i y_i^*$, given vector $Y = (y_1 \quad y_2 \quad ... \quad y_n)^T$ and basis $p_1 = (x_1 \quad x_2 \quad ... \quad x_n)$ and $p_2 = (1 \quad 1 \quad ... \quad 1)$, we should find paramter $c = (\alpha \quad \beta)^T$ s.t. the linear combination $\hat{Y} = \beta p_1 + \alpha p_2$ is the best estimation of Y in terms of its square error $\|e\|^2 = \sum_i (y_i - \alpha - \beta x_i)^2$. Then linear regression can be easily settled by solving the following normal equation

$$A^\dagger A c = A^\dagger Y \Rightarrow c = (A^\dagger A)^{-1} A^\dagger Y. \tag{22}$$

We can further use paramters c to predict data y from certain x:

$$y_{\text{predict}} = (1 \quad x) \left[(A^\dagger A)^{-1} A^\dagger Y \right]. \tag{23}$$

3.3.3 Least square filtering

We will briefly go through this. Consider a filter design problem. Given a desired output $y[n]$ and input $x[n]$, how to find a filter $h[n]$ such that error

$$e[n] = (x * h)[n] - y[n] \tag{24}$$

is minimized in terms of $\sum_n e[n]^2$. Suppose $h[n]$ has limited length, so $y[n] = (x*h)[n] + e[n] = y[n] = \sum_i x[i]h[n-i] + e[n]$ can be written in the form of

$$y = Ah + e, \tag{25}$$

where A is decided by $x[n]$. We can see that it is able to transform filtering problem into a linear projection problem thanks to the linearity of filtering.

4 Dual approximation problem

At last we are going to mention a dual problem of projection approximation. Let us consider the following problem.

Definition 4.1. *We call the following a dual problem of projection approximation. Let S be Hilbert space and $\{y_i \in S\}$ is a linear independent basis spanning $M = \{y_i\}$. We have an unkown vector x, such that $\langle x, y_i \rangle = b_i$. Our goal is to find a x with minimized $\|x\|^2$.*

We will see in a second that this problem is actually a projection problem. Suppose $x = \sum_i c_i y_i + m^\perp$, where $\sum_i c_i y_i \in M$ and $m^\perp \in M^\perp$. We call $m^\perp$ a null solution, and it has no effect on inner product $\langle x, y_i \rangle$. So $m^\perp$ can be picked up randomly. Besides, because $\{y_i\}$ is linear independent, so $\{c_i\}$ is unique determined by $\{b_i\}$. Now we are going to explain this problem as a projection approximation. Vector $\sum_i c_i y_i$ is the vector we desire, and $m^\perp$ is the projection vector, and x is the projection error. To minimize error, we require $x \perp M^\perp$. Thus $m^\perp = 0$ and we have $x = \sum_i c_i y_i$. Use condition $\langle x, y_i \rangle = b_i$, we have the following equation:

$$\sum_i c_i \langle y_i, y_j \rangle = \langle x, y_j \rangle. \tag{26}$$

Define grammian $R_{i,j} = \langle y_j, y_i \rangle$ and $b_j = \langle x, y_j \rangle$, we can rewrite it in a matrix form

$$Rc = b. \tag{27}$$

We find this problem has a exact parallel form with projection approximation (though the parameters have differernt meanings). So this normal equation actually represents two different but dual problems.

The importance of this dual problem is because it is equivalent to a matrix equation. Consider matrix equation

$$Ax = b. \tag{28}$$

Note that A is not necessarilly square matrix. So we cannot simply find an inverse of A. But if we write A in the way of

$$A_{i,j} = (y_i)_j, \tag{29}$$

where y_i is a vector. Then $(Ax)_i = \sum y_i x_i = \langle x, y \rangle$. So $Ax = b$ is equivalent to

$$\langle x, y_i \rangle = b_i. \tag{30}$$

This is exactly the dual problem we gave above (if $\{y_i\}$ is linear independent). Then if we ask for the solution x with minimum $\|x\|^2$, then it is equivalent to solve the normal equation

$$Rc = (AA^\dagger)c = b. \tag{31}$$

Since $\{y_i\}$ is linear independent, R has inverse $R^{-1} = (AA^\dagger)^{-1}$. The solution x with minimized $\|x\|^2$ is

$$x = A^\dagger c = A^\dagger (AA^\dagger)^{-1} c. \tag{32}$$

5 Conclusion

Many linear optimization problem can be seen as a projection approximation problem, where we can apply projection theorem as a general method. Then it always leads to a normal equation, though in different circumstances the equation has different meaning. Thanks to the generality of projection theorem, we are able to deal with many optimization problems in a unified frame.