1. Least Squares, The Orthogonal projection Lemma, and the Discrete-Time Kalman Filter

* Review Chapter.3 : the conditional expectation

In Chapter3.4 the discrete-time Kalman filter is considered. The summary is following

1. Problem formulation : a dynamic system is considered as
2. Find the mean minimum mean square error estimator given

i.e.,

1. The solution

Based on the conditional expectation which is a Gaussian, and recursive method, it is divided into the prediction and the estimation(correction the error between the prediction and the measurement).

-. Prediction

And its covariance

-. Estimation

And the error covariance

1. Analysis or comments

* In chapter 3, the derivation is based on the fact that the conditional expectation is a Gaussian due to the initial condition of states and the noise being Gaussians
* If they are Gaussians, what is the best estimator?
* The derivation is the conditional expectation in Chapter 3, whereas in chapter.4 the other method is studying, such as the least square estimation, which is originally derived by Kalman
  1. Least Squares
* The problem :
* A sequential measurements , which is

Minimize the estimator of to satisfy the cost function as

%%% Kim’s comment: least square estimator

Example: Find the straight line such that the error is minimized

%%%

* Least Square - Generalization:

The least square cost of function is

To find to minimize the cost function, we need the derivative of the function with respect to the vector

%%% Vector Calculus

-example\_1

And

-example\_2

%%%

* The vector derivative of the cost function

First

Then take the derivative at both sides

For the optimality, the gradient should be zero,

Imply to

Or using “**Calculus of Variation**” (You may learn at the next semester in “optimal control”)

* 🡪

%%% Kim’s comment

The linear equation:

1. The equation has the unique solution
2. The equation has infinite many solution
3. The equation has non solution 🡪 pseudo inverse (satisfying the min least square errors)

* If exists,
* **Ex. 4.5**

1. Condition:

The position of a vehicle are measured as

1. Problem : estimate initial position, velocity, acceleration, using LSE

-original data

-measured data

The

1. Two models are considered:

Model\_1: , find the estimator of

Model\_2 : find the estimator of

1. Results.

-.The estimators of Model\_1 are good. The estimators of Model\_2 are bad.

-.If the measurement of noise is larger(the variance), the estimator is not good in both case, Model\_1 and Model\_2. Here, if the information of the noise is not considered. If it is considered, then it is the equivalent to the chapter.3 solved using the conditional expectation.

* Def. 4.1 A matrix, , is called an orthonormal matrix if

%%% Kim’s

Ex.

1. Orthogonal : each column vectors of are orthogonal,
2. Orthonormal:
3. Orthonormal matrix is a square matrix, #of rows = #of columns

%%%

* **Theorem 4.2 (singular value decomposition)** Let be a real matrix. There exists an orthonormal matrix and an orthonormal matrix and an real matrix,such that

is of the form

where is a real and

The real **scalars**  are called the singular values of

This is better to understand as a definition of the singular value decomposition

%%% Kim’s comment: eigenvalue / vector

For a square matrix ,

where

Hence :

1. Eigenvalue : the matrix should be a square, but SVD is not needed.
2. Eigenvalue is different from the singular value even if the given matrix is a square.
3. Some square matrices (i.e. positive definite) they are closely related.

i.e., Given a matrix A, the square root of eigenvalues of is the sigma value

1. Example

* Interpretation of SVD

Now

* For input , the only will be effective ,
* For output the only ( with a scale will be measured.
* For the biggest the input is the biggest output in the direction of
* For the smallest the input is the smallest output in the direction of , i.e., it may be the least observable in the output
  1. The orthogonal projection lemma

%% Gram-Schmidt ortho-normalization

* Problem: In a n-dimensional vector space ,

Given any independent set of vectors

which is a basis of a Vector space. Find a basis which is orthonormal to

each other.

* Solution - Gram-Schmidt procedure:

1. Select any vector by simply normalizing to be the first basis vector
2. For the next basis vector, see the following drawing as an example

Here is orthogonal to , subtracting from that part of the that lies along

since

To be normalized,

1. For the third elements, calculate

To prove the orthogonality, take inner product at both sides

And

Hence is orthogonal to

We may prove orthogonality using linear property of the inner product as

1. For the next element, proceed as the previous procedure..
2. The general formula is

%%% HA\_7\_1

1. Prove
2. Here the given independent vector set’s elements number is n. So the result orthogonal vector set’s element number is “n”. If the original given set of vector is not independent, what is the number of the result orthogonal vector set’s element number? Prove it.

%%%

* Hilbert space

1. If is a Hilbert space and if then can be described by a set of basis vector, , : the number of basis may be infinite.
2. Euclidean space: the real space(not abstract space), the number of basis( in 3D) is finite.

* Inner product (Wiki: <https://en.wikipedia.org/wiki/Inner_product_space> )

Assume is a vector space. Define a mapping (an operator) ,

i.e., for any vector

If the mapping satisfies

i). Linearity

and

ii). Positive semi definiteness

And

Then the mapping( or operator) is called an inner product.

Ex. Inner product.

i).Dot product:

then 🡪 this operator satisfies the linearity and semi positivity.

* is a real inner product space, and then

1. iff

Norm: , measure, a notion of distance / length in Hilbert space

%% kim’s comment

What is Hilbert Space?

1. The number of elements in basis is as much as infinite.
2. Euclidean Space is a subspace of Hilbert Space.
3. Special case: Let a continuous function space on [0, T]

Then

A basis =

where the inner product is defined as: %%%

%%% Kim’s comment : Inner product in probability space:

Two random variables, , the expectation of two RVs is an inner product.

i.e.,

In this case iff

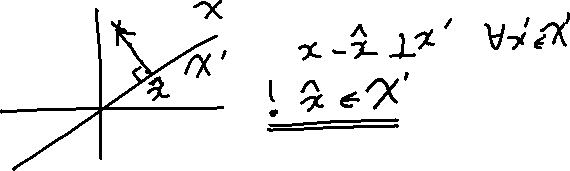
So the RVs ,

%%%

* **Lemma 4.7 (Orthogonal Projection Lemma**) : Let be a Hilbert space with and let be a subspace of . Then there exists a unique vector such that

Iff

* **Implication of the Lemma: the best possible estimate of a function, or a vector, is the one that is orthogonal to its estimation error.**



* 1. Extensions of Least Squares Theory(skip)
  2. Non-linear Least Squares: Newton-Gauss Iteration
* Problem

Find the Least Square Estimator for with the measurement

%%% Kim’s comment

In linear case, for example, based on the governing equation

We may measure at which are based on the linear equation. Here

Now the governing equation is

Then the equation is not linear, i.e., we can not have a matrix from as (a). %%%%

In non linear system, we may use an first order approximation as follows

1. The first-order approximation

At the initial guess,,

The cost function with the linearized approximation

where

Here

Eq.(4.22) is a linear form which is the same form of linear least square, so the LSE is

Then the initial estimator based on LSE is

where

Now the first non-linear estimator is assumed following the first estimator base on LSE

1. The second non-linear estimator Using the first estimator, we may continue

where

1. Now in general up to Newton-Gauss iteration

Here

%%% Kim’s comment

* The pseudo inverse
* The matrix is called the Jacobian whose row vector is

Here the jacobian is defined , in general, as , given

Here the transpose of the Jacobian is the gradient of the variables.

* In linear case, at the time , the Jacobain is

which is independent of the variable a and b.

However, in the example of non-linear

which is dependent of a and b. Also, for

* As the error

Also is changed for

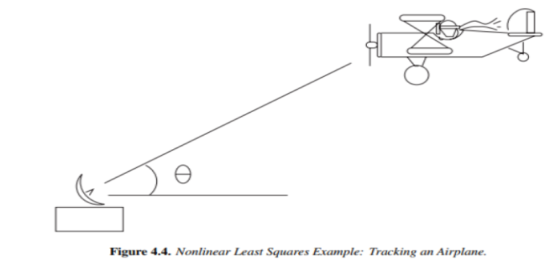
%%% kim’s comment

with the measurement , as

Find an estimate

1. initial guess :
   1. Calculate at the first guess

* Example. 4.9 (Tracking an airplane)



1. Problem :

* An airplane travels at a constant speed in the vertical plane
* The measurement is only the sight(angle) of a radar station
* To estimate the airplane’s position

1. Modeling

* System dynamics
* Measurements: the angle (line-of-sight)

Here, as in the linear case, we estimate

1. Process;

3.1) Here the measurement

is nonlinear w.r.t , to apply Newton-Gauss method, at every measurement point, we need linearize the measurement w.r.t.

To do linearization, the first order approximation

,

* 1. Deriving the Kalman Filter via the Orthogonal Projection Lemma
* Problem :

Where the measurement noise are assumed to be zero-mean, uncorrelated but **not necessarily Gaussian.**

1. Measurement space: Define to be a vector space , a subspace of the measurement space, as
2. The cost function:

* **Derivation : projection**
* We know the best estimator should satisfy the orthogonal projection lemma

Hence the best estimator (in this case least square sense), satisfy

1. The first step
2. Apply orthogonal Projection: it is necessary

Where is any arbitrary linear function of , which is **in the measurement space.**

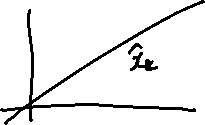
1. The orthogonal basis representation of

Let define an orthogonal basis of as

Which is

From the orthogonal projection lemma, is optimal iff

Or



which implies

%% Kim’s comment : Basis representation :

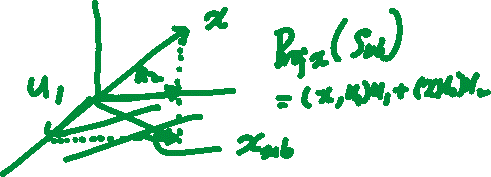
An orthonormal basis = {(1,1),(1,-1)}. Let find x = a(1,1) + b(1,-1)

* Solution: %%

%% Kim’s comment : Projection of on the subspace

Let Define a subspace , is a subspace of , whose a basis . Then the projection of on the subspace is

Let , then



%%%

* Hilbert Space introduced.(4.29.1)

Since , A vector presentation of is , since is the i-th element of the orthogonal basis,

1. Recursive way concerning to the system
2. The innovation process:

What is the which is a basis of the measurement space of ?

We will pick as

which is orthogonal to

Proof: Applying the orthogonal projection lemma, it is

Multiply by

Multiply by

Therefore



is orthogonal to which is a valid choice

1. Kalman Gain



In (4.33)

We need the proportional gain. To begin with

From (4.32.1)

Define And Then

.

where

and

We then get

1. The error covariance

* Implications:

The error is orthogonal to the measurements

Since The error is orthogonal to the

1. Define the innovation as

Then

And

* Summary of Ch.4

Using Orthogonal projection Lemma, find the Kalman gain

1. The cost function:
2. Using the Lemma
3. Guess , the orthonormal basis of the measurement space
4. **Find , which is the Kalman Gain**

Where

1. For calculation of , we need an iterative method. As

Given initial conditions as

+ Initial state conditions (mean, variance):

+ Initial system parameters:

+ Initial noise parameters:

Step\_1: Prediction Step

+ mean:

+ variance:

Step\_2: Estimation (After measurement)

+ mean**:**

**+** variance**:**