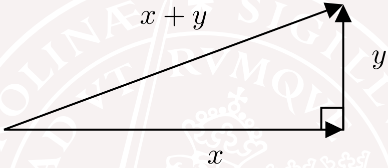
Lecture 6

- Hilbert Spaces
- Least squares problems - underdetermined
  - Measures of controllability
- Least squares problems - overdetermined
  - Measures of observability
- Example: Function approximation

Motivation: Least Squares Minimization



$$||x + y||^2 \;=\; ||x||^2 \;+\; 2x^T y \;+\; ||y||^2$$

If x and y are orthogonal, i.e. $x^T y = 0$, then

$$||x + y||^2 \;=\; ||x||^2 \;+\; ||y||^2$$

Minimization of this over y is trivial, giving $y = 0$ and

$$\min_y ||x + y||^2 \;=\; ||x||^2$$

Hilbert Spaces

Want to generalize to a situation were we optimize over for example "all possible control signals $u([0,T])$".

How should we generalize orthogonality $u^T v = 0$, and norm $\|u\|$?

Hilbert space = "Complete, normed vector space with scalar product"

Scalar product $\langle u, v \rangle$ satisfying some natural axioms

Norm $\|u\|^2 = \langle u, u \rangle$

Example: $L_2[0, t_1]$, square-integrable functions, with scalar product

$$\langle y, w \rangle = \int_0^{t_1} y^T(t) w(t) dt$$

Linear operators and Hilbert spaces

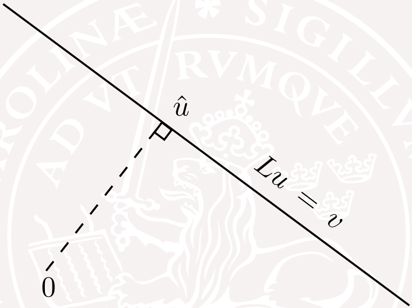
Useful theory: Linear operators in infinite-dimensional vector spaces, scalar product $\langle u, v \rangle$, "orthogonal" means that $\langle u, v \rangle = 0$.

This theory has many applications, not only in control and signal processing.

Don't have time here to present all mathematical details, only some intuition and the resulting formulas for the optimal solution.

For more detail, see Lecture 6 in the PhD course Linear System theory
`www.control.lth.se/Education/DoctorateProgram/linear-systems.html`

Least squares problem I



Given linear operator L and vector v:
Minimize $\|u\|$ under the constraint $Lu = v$.

The operator L is “short and fat”: More variables than equations.
Typically many solutions u, want the shortest one.
Notice the right angle in the picture

Linear operators and Hilbert spaces

If L is a matrix and u and v are vectors this is an easy matrix problem: The solution, if the columns of L span the full space, is

$$\hat{u} \;=\; L^T(LL^T)^{-1}v \tag{1}$$

The matrix $L^\dagger = L^T(LL^T)^{-1}$ is called the pseudo-inverse of L.

We want to generalize this problem and solution to a situation were we optimize over for example "all possible control signals $u([0,T])$".

A proof of (1) follows from the more general result below

Linear operators and adjoints

Given a linear operator L from a Hilbert space H_1 to another H_2. The adjoint L^* is an operator from H_2 to H_1 defined by the relation

$$\langle Lu_1, u_2 \rangle = \langle u_1, L^* u_2 \rangle \tag{2}$$

for all $u_1 \in H_1, u_2 \in H_2$.

This generalizes the matrix transpose in the finite dimensional case

Note that the scalar products in (2) are in different spaces

Example

The operator

$$Lx_0 = Ce^{At}x_0, \quad t \in [0, t_1]$$

maps x_0 in $H_1 = \mathcal{R}^n$ to the function $Ce^{At}x_0$ in $H_2 = L_2[0, T]$

Claim: The adjoint L^*, mapping functions to vectors, is given by

$$L^*y = \int_0^{t_1} e^{A^Tt}C^Ty(t)dt$$

Proof: We just need to verify that (2) holds:

$$\langle Lx_0, y\rangle = \int_0^{t_1}(Ce^{At}x_0)^Ty(t)dt = x_0^T\int_0^{t_1} e^{A^Tt}C^Ty(t)dt = \langle x_0, L^*y\rangle.$$

The left is a scalar product in L_2, the right in $\mathcal{R}^n$

Another example

The operator ($M = L^*$ after some variable renaming)

$$Mu = \int_0^{t_1} e^{At} B u(t) dt$$

maps the function $u(t)$ in $H_1 = L_2[0, T]$ to a vector in $H_2 = \mathcal{R}^n$.

Claim: The adjoint M^*, mapping vectors to functions, is given by

$$M^* x = B^T e^{A^T t} x$$

Proof: We again just need to verify that (2) holds:

$$\langle Mu, x \rangle = \left[\int_0^{t_1} e^{At} B u(t) dt \right]^T x = \int_0^{t_1} u^T B^T e^{A^T t} x dt = \langle u, M^* x \rangle.$$

The claim alternatively follows from the general rule $(L^*)^* = L$

Least squares problem I

Minimize $\|u\|$ under the constraint $Lu = v$.

Claim 1: An optimal $\hat{u}$ must satisfy $L\hat{u} = v$ and ("orthogonality")

$$0 = \langle \hat{u}, u - \hat{u} \rangle \text{ for all } u \text{ with } Lu = v \qquad \text{(OC1)}$$

Proof: Necessity of the orthogonality condition follows from derivating

$$f(t) = \|\hat{u} + t(u - \hat{u})\|^2 = \|\hat{u}\|^2 + 2t\langle \hat{u}, u - \hat{u} \rangle + t^2\|u - \hat{u}\|^2$$

and setting $t = 0$.

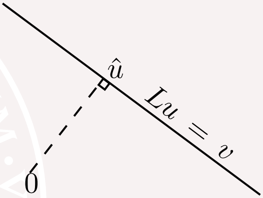
If $L\hat{u} = v$ and (OC1) holds, then $\hat{u}$ is optimal and unique, since

$$\|u\|^2 = \|\hat{u}\|^2 + \|u - \hat{u}\|^2.$$

Solution to LS 1

Claim 2 If LL^* is invertible then the solution to LS1 is

$$\hat{u} = L^*(LL^*)^{-1}v$$



Proof: Obvious that $L\hat{u} = v$. Furthermore

$$\begin{aligned}\langle \hat{u}, u - \hat{u}\rangle &= \langle L^*(LL^*)^{-1}v, u - L^*(LL^*)^{-1}v\rangle = \\ &= \langle (LL^*)^{-1}v, Lu - LL^*(LL^*)^{-1}v\rangle \\ &= \langle (LL^*)^{-1}v, Lu - v\rangle = 0\end{aligned}$$

So the orthogonality condition is satisfied.

Note also that $\|\hat{u}\|^2 = \langle v, (LL^*)^{-1}v\rangle$

Let's apply LS1 to the problem of controllability

The controllability problem

On the previous lecture we saw that finding a control $u(t)$ which moves $x(0) = x_0$ to $x(t_1) = 0$ gives the linear equation

$$\int_0^{t_1} e^{-At} Bu(t)dt = -x_0$$

Introduce therefore $v = -x_0$ and the operator

$$Lu = \int_0^{t_1} e^{-At} Bu(t)dt$$

which maps a function $u(t) \in L_2^m[0, t_1]$ to $\mathcal{R}^n$. Use scalar product

$$\langle u, v \rangle = \int_0^{t_1} u^T(t)v(t)dt$$

Measure of controllability

$$Lu = \int_0^{t_1} e^{-At} Bu(t)dt = v$$

$$L^* w = \left[e^{-At} B\right]^T w$$

$$W := LL^* = \int_0^{t_1} e^{-At} BB^T e^{-A^T t} dt$$

The problem to control $\dot{x} = Ax + Bu$ from $x(0) = -v$ to $x(t_1) = 0$ with minimal cost $\|u\|^2 = \int_0^{t_1} \|u(t)\|^2 dt$ hence has solution

$$\hat{u}(t) = L^*(LL^*)^{-1} v = -B^T e^{-A^T t} W^{-1} x_0$$

and the minimal squared cost equals

$$\|\hat{u}\|^2 = x_0^T (LL^*)^{-1} x_0 = x_0^T W^{-1} x_0.$$

Another Gramian

The slightly different matrix

$$W_r = \int_0^{t_1} e^{A\tau} B B^T e^{A^T \tau} d\tau$$

is called the reachability Gramian.

It measures the cost of going from $x(0) = 0$ to $x(t_1) = x_1$

The squared cost is $x_1^T W_r^{-1} x_1$

The smallest eigenvalue of W_r is a measure of controllability, since $1/\lambda_{\min}(W_r)$ is the control signal (squared) norm that is needed to reach all states having norm one.

For the case $t_1 = \infty$ and A asymptotically stable, one can calculate W_r from the Lyapunov equation (`Wr=lyap(A,B*B')` in matlab)

$$W_r A^T + A W_r + B B^T = 0.$$

Example: Gramian for trailer

$$A = \begin{bmatrix} -1 & 0 \\ 1 & -1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad e^{At} = \begin{bmatrix} e^{-t} & 0 \\ te^{-t} & e^{-t} \end{bmatrix}$$

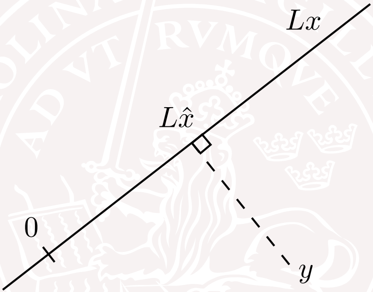
$$\begin{aligned} W_r &= \int_0^{t_1} \begin{bmatrix} e^{-t} \\ te^{-t} \end{bmatrix} \begin{bmatrix} e^{-t} \\ te^{-t} \end{bmatrix}^T dt \\ &= \frac{1}{4} \begin{bmatrix} 2 - 2e^{-2t_1} & 1 - (2t_1 + 1)e^{-2t_1} \\ 1 - (2t_1 + 1)e^{-2t_1} & 1 - (2t_1^2 + 2t_1 + 1)e^{-2t_1} \end{bmatrix} \end{aligned}$$

For $t_1 = \infty$ we get

$$W_r = \begin{bmatrix} 1/2 & 1/4 \\ 1/4 & 1/4 \end{bmatrix}$$

with eigenvalues 0.65 and 0.096.

Least squares problem II



Given L and y, find x that minimizes the error $\|Lx - y\|$

L is “tall and thin”: More equations than variables

Notice the right angle !

Least squares problems II

Minimize $\|Lx - y\|$ with respect to x.

Claim: An optimal $\hat{x}$ must satisfy

$$0 = \langle Lw, L\hat{x} - y \rangle \text{ for all } w \qquad \text{(OC2)}$$

Proof: Derivate $\|L(\hat{x} + tw) - y\|^2$ with respect to t.
Note that (OC2) is equivalent to

$$L^* L\hat{x} = L^* y$$

If $L^* L$ is invertible, then the solution is unique and given by

$$\hat{x} = (L^* L)^{-1} L^* y$$

Observability

$$\begin{cases} \dfrac{dx}{dt} = Ax, \quad x(0) = x_0 \\ \quad y = Cx \end{cases}$$

The system is observable if x_0 uniquely can be determined from $y_{[0,t_1]}$.

$$y(t) = Ce^{At}x_0 = (Lx_0)(t), \quad y = Lx_0$$
$$L : \mathbf{R}^n \to \mathbf{L}_2^p[0, t_1]$$

The operator L now maps x_0 to y, i.e. from an n-dimensional space to a space of functions

Measure of observability

If $y(t) = Lx_0 + e(t)$, i.e. if measurements are disturbed by noise $e(t)$, then typically no x_0 can be found solving $y = Lx_0$ perfectly
Least squares solution:

$$\min_{x_0} ||y - Lx_0||$$

$$W_o = L^*L = \int_0^{t_1} e^{A^Tt}C^TCe^{At}dt$$

$$\hat{x}_0 = (L^*L)^{-1}L^*y = W_o^{-1}\int_0^{t_1} e^{A^Tt}C^Ty(t)dt$$

Since $L\hat{x}_0 = Lx_0 + e$ the estimation error $\tilde{x}_0 = x_0 - \hat{x}_0$ satisfies

$$\tilde{x}_0^T L^* L\tilde{x}_0 = \|e\|^2$$

The smallest eigenvalue to the *observability gramian* $W_o = L^*L$ gives a measure of observability. If it is close to zero, then small e can give large $\tilde{x}_0$ (bad).

Other example: Function approximation

Choose the real numbers a_0, a_1, a_2 to minimize $\int_0^1 |e^t - a_0 - a_1 t - a_2 t^2|^2 dt$

Solution:

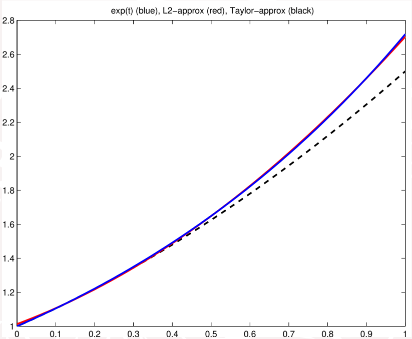
$$x = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}, \qquad Lx = \begin{bmatrix} 1 & t & t^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}, \qquad y(t) = e^t$$

$$L^* y = \int_0^1 \begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix} e^t dt = \begin{bmatrix} e-1 \\ 1 \\ e-2 \end{bmatrix}$$

$$L^* L = \int_0^1 \begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix} \begin{bmatrix} 1 & t & t^2 \end{bmatrix} dt = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}$$

$$\widehat{x} = (L^* L)^{-1} L^* v = \begin{bmatrix} 1.013 \\ 0.851 \\ 0.839 \end{bmatrix}$$

Example: Function approximation



Notice that the least squares approximation

$$e^t \approx 1.013 + 0.851t + 0.839t^2$$

is much better than the Taylor approximation (dashed)

$$e^t \approx 1 + t + 0.5t^2$$

Lecture 6

- Hilbert Spaces
- Least squares problem, under-determined
- Measures of controllability
- Least squares problem, over-determined
- Measures of observability
- Example: Function approximation

End of the course