Lecture 6

- Least squares problems
- Adjoint operators

Review: Least Squares Solution to Linear Equations (I)

Consider a system of linear equations

$$Ax = b, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$$

with $m \geq n$ and $\text{rank}(A) = n$ (Tall A—more rows than columns, or more equations than unknowns).

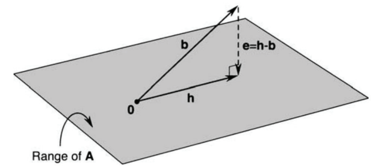
If $b \notin \text{range}(A)$ then the linear system is inconsistent, i.e., no solution exists.

Find x that minimizes $\|Ax - b\|^2$—least squares solution

Review: Least Squares Solution to Linear Equations (I)

Least squares solution to the inconsistent linear equation $Ax = b$ is given by the solution to $A^\top Ax = A^\top b$; i.e., $x_{\text{ls}} = (A^\top A)^{-1}A^\top b$.

Geometric interpretation:



Orthogonal projection of b on the subspace range(A).

Review: Least Norm Solution (II)

Consider a system of linear equations

$$Ax = b, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$$

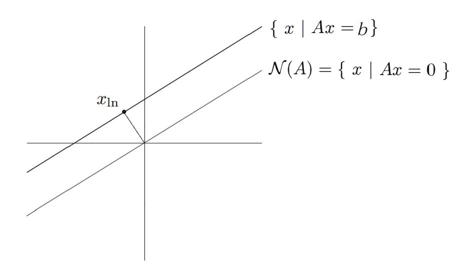
with $m \leq n$ and rank$(A) = m$ (Fat A—more columns than rows, or more variables than equations).

- There exist an infinite number of solutions to this linear equation—(Underdetermined)
- There is only one solution that is closest to the origin; i.e., a solution to $Ax = b$ with least norm $\|x\|$.

Review: Least Norm Solution (II)

Least norm solution to the underdetermined linear equation $Ax = b$ is given $x_{\text{ln}} = A^\top (AA^\top)^{-1} b$.

Geometric interpretation:

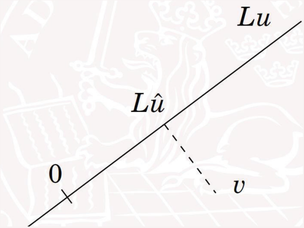


- orthogonality condition: $x_{\text{ln}} \perp \text{Null}(A)$
- projection interpretation: x_{ln} is projection of 0 on solution set $\{x|Ax = b\}$.

Least Squares Problems I

Given L and v, minimize $|Lu - v|$ with respect to u.

"Tall L, more equations than variables"

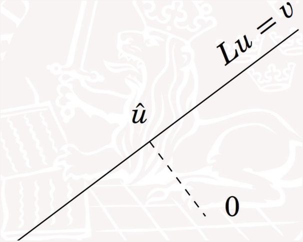


Note the orthogonality in the picture!

Least Squares Problems II

Given L and v, minimize $|u|$ under the constraint $Lu = v$.

"Fat L, more variables than equations"



Note the orthogonality in the picture!

Vector Space and Inner Products

A vector space V is a generalisation of R^n and is defined by 'vectors' and 'scalars' satisfying some standard rules, e.g

- addition of vectors, $v_1 + v_2 = v_2 + v_1$, $v_1 + (v_2 + v_3) = (v_1 + v_2) + v_3$
- multiplication with scalars $(\lambda_1 \lambda_2)v = \lambda_1(\lambda_2 v)$

(there are more rules). Scalar field could be e.g. R och C

A scalar product $< V, V > \to R$ is a generalisation of $y^* x$ satisfying natural linearity rules and $< v, v > \in (0, \infty)$ forall nonzero v.

Note that for complex scalars

$$< \lambda_1 v_1, \lambda_2 v_2 > = \bar{\lambda}_1 \lambda_2 < v_1, v_2 >$$

Examples and Orthogonality

Finite-dimensional vector space:

$R^n, \quad < y, x > = y^* x$ or $< y, x > = y^* Q x, \; Q > 0$

$R^{n \times m}, \quad < Y, X > = \mathrm{tr}(Y^* X)$

Infinite-dimensional vector space:

$l_2, \quad < y, x > = \sum_{k=1}^{\infty} y_k^* x_k$

$L_2[a, b], \quad < y(t), x(t) > = \int_a^b y^*(t) x(t) dt$

$L_{2,w}[a, b], \quad < y(t), x(t) > = \int_a^b y^*(t) x(t) w(t) dt, \; w > 0$

We will say that x and y are *orthogonal* if

$$< x, y > = 0$$

Vectors orthogonal to a subspace S will be denoted by $S^\perp$
(Orthogonal complement)

Examples and Orthogonality

Finite-dimensional vector space:

$R^n, \quad < y, x > = y^* x$ or $< y, x > = y^* Q x, \; Q > 0$

$R^{n \times m}, \quad < Y, X > = \mathrm{tr}(Y^* X)$

Infinite-dimensional vector space:

$l_2, \quad < y, x > = \sum_{k=1}^{\infty} y_k^* x_k$

$L_2[a, b], \quad < y(t), x(t) > = \int_a^b y^*(t) x(t) dt$

$L_{2,w}[a, b], \quad < y(t), x(t) > = \int_a^b y^*(t) x(t) w(t) dt, \; w > 0$

We will say that x and y are *orthogonal* if

$$< x, y > = 0$$

Vectors orthogonal to a subspace S will be denoted by $S^{\perp}$
(Orthogonal complement)

Examples and Orthogonality

Finite-dimensional vector space:

$R^n, \quad <y, x> = y^* x$ or $<y, x> = y^* Q x, \; Q > 0$

$R^{n\times m}, \quad <Y, X> = \mathrm{tr}(Y^* X)$

Infinite-dimensional vector space:

$l_2, \quad <y, x> = \sum_{k=1}^{\infty} y_k^* x_k$

$L_2[a, b], \quad <y(t), x(t)> = \int_a^b y^*(t) x(t) dt$

$L_{2,w}[a, b], \quad <y(t), x(t)> = \int_a^b y^*(t) x(t) w(t) dt, \; w > 0$

We will say that x and y are *orthogonal* if

$$<x, y> = 0$$

Vectors orthogonal to a subspace S will be denoted by $S^\perp$ (Orthogonal complement)

Example: Matrix Adjoint

Let $L : X \to Y$ be a bounded linear operator. The *adjoint operator* $L^* : Y \to X$ is defined by the identity

$$< y, Lx > \quad = \quad < L^* y, x >$$

for $x \in X$, $y \in Y$.

From the equalities

$$< y, Lx > = y^* Lx = (L^* y)^* x = < L^* y, x >$$

we see that the adjoint of a matrix is given by its conjugate transpose.

Example: Adjoint Transition Matrix

If $L : \mathbf{R}^n \to \mathbf{L}_2^m[0,\infty)$ is defined by

$$(Lx_0)(t) = C(t)\Phi(t,0)x_0, \quad x_0 \in \mathbf{R}^n$$

then the adjoint $L^* : \mathbf{L}_2^m[0,\infty) \to \mathbf{R}^n$ is given by

$$L^* y = \int_0^\infty \Phi(t,0)^T C(t)^T y(t) dt$$

Proof:

$$\begin{aligned} < y, Lx_0 > &= \int_0^\infty y(t)^T C(t)\Phi(t,0)x_0 dt \\ &= \left(\int_0^\infty \Phi(t,0)^T C(t)^T y(t) dt \right)^T x_0 \\ &= < L^* y, x_0 > \end{aligned}$$

Exercise

Define instead $L : \mathbf{L}_2^m[0,\infty) \to \mathbf{L}_2^m[0,\infty)$ by

$$(Lu)(t) \quad = \quad \int_0^t \Phi(t,s)u(s)ds$$

What is the adjoint L^*?

Hint: $< y, Lu > = < L^*y, u >$

$\boxed{\text{The audience is thinking}}$

Adjoint Equation

If $\Phi(t, t_0)$ is the transition matrix for

$$\dot{x}(t) = A(t)x(t)$$

then $\Phi(t_0, t)^T$ is the transition matrix for

$$\dot{z}(t) = -A(t)^T z(t)$$

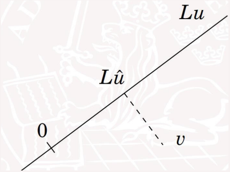
The relation can be written

$$[\Phi_A(s,t)]^* = \Phi_{-A^T}(t,s)$$

Proof: Exercise

Least Squares Problem I

Minimize $|Lu - v|$ with respect to u.



Solution: Any $\hat{u}$ satisfying the Orthogonality Property

$$0 = < Lx, L\hat{u} - v > \text{ for all } x \tag{OP1}$$

Or equivalently

$$L^* L\hat{u} \;\; = \;\; L^* v$$

Application: Fewer control signals than objectives

Proof - Completion of Squares

Assume first $\hat{u}$ satisfies OP1. Then with $x = u - \hat{u}$ we get

$$\begin{aligned} |Lu - v|^2 &= < L\hat{u} - v + Lx, L\hat{u} - v + Lx > \\ &= < L\hat{u} - v, L\hat{u} - v > + < Lx, Lx > \\ &= \ |L\hat{u} - v|^2 + |Lx|^2 \\ &\geq |L\hat{u} - v|^2 \end{aligned}$$

Therefore $\hat{u}$ is optimal (might be non-unique).

Assume instead that $\hat{u}$ is optimal. Since for any x

$$|L(\hat{u} + \epsilon x) - v|^2 =$$
$$< L\hat{u} - v, L\hat{u} - v > \ +2\epsilon < Lx, L\hat{u} - v > \ +\epsilon^2 < Lx, Lx >$$

should be min for $\epsilon = 0$ we see $< Lx, L\hat{u} - v > = 0$, i.e OP1.

Example: Estimating Initial State

Define L : $\mathbf{R}^n \to \mathbf{L}_2^m[t_0, t_1]$ by

$$(Lx_0)(t) \quad = \quad C(t)\Phi(t_1, t_0)x_0, \quad x_0 \in \mathbf{R}^n$$

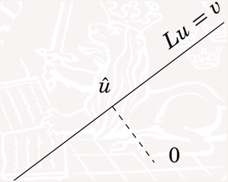
Problem:
Given an output measurement $y(t)$ for $t \in [t_0, t_1]$, find the value of x_0 that minimizes $|Lx_0 - y|$.

Solution:
We calculated $L^*y = \int_{t_0}^{t_1} \Phi(t, t_0)^T C(t)^T y(t)dt$ above.
Use of OP1 formula gives

$$\begin{aligned} x_0 &= (L^*L)^{-1}L^*y \\ &= \left(\int_{t_0}^{t_1} \Phi(t, t_0)^T C^T C\Phi(t, t_0)dt\right)^{-1} \int_{t_0}^{t_1} \Phi(t, t_0)^T C^T y(t)dt \end{aligned}$$

Least Squares Problem II

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



Solution: Any $\hat{u}$ satisfying $L\hat{u} = v$ and the Orthogonality Property

$$0 = < \hat{u}, \hat{u} - u > \text{ for all } u \text{ with } Lu = v \qquad \text{(OP2)}$$

Or, if LL^* invertible, equivalently

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

Application: Reach certain state with minimal cost

Proof - Completion of Squares

Assume a candidate $\hat{u}$ satisfies OP2. Then

$$< u, u > \; = < \hat{u}, \hat{u} > + < u - \hat{u}, u - \hat{u} > \; \geq \; < \hat{u}, \hat{u} >$$

for all u satisfying $Lu = v$. Hence $\hat{u}$ is optimal (and unique).

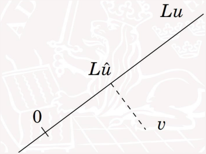
Necessity of OP2: As above, study $\|\hat{u} + \epsilon(u - \hat{u})\|^2$ near $\epsilon = 0$

If LL^* invertible then $\hat{u} = L^*(LL^*)^{-1}v$ satisifies both $L\hat{u} = v$ (obvious) and OP2:

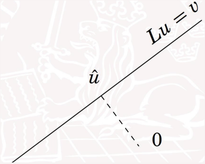
$$\begin{aligned} < \hat{u}, \hat{u} - u > &= \; < L^*(LL^*)^{-1}v, L^*(LL^*)^{-1}v - u > \\ &= \; < (LL^*)^{-1}v, v - Lu > \; = 0 \end{aligned}$$

Remarks - Technical Details

In the first problem, the solution $\hat{u}$ might be nonunique if L^*L is not invertible. For example when $L = 0$

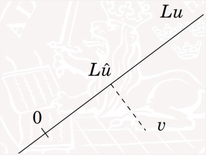


In the second problem, if LL^* is not invertible, the equation $Lu = v$ might be unsolvable, the solution can also be non-unique. But if $LL^*x = v$ is solvable, then $\hat{u} = L^*x$ is optimal

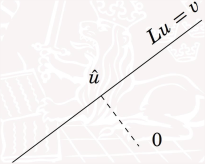


Remarks - Technical Details

In the first problem, the solution $\hat{u}$ might be nonunique if L^*L is not invertible. For example when $L = 0$



In the second problem, if LL^* is not invertible, the equation $Lu = v$ might be unsolvable, the solution can also be non-unique. But if $LL^*x = v$ is solvable, then $\hat{u} = L^*x$ is optimal



Example

Find $u(t) = \begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix}$ with min 2-norm so $\int_0^1 u_1(t) + tu_2(t)dt = 4$

$\boxed{\text{The audience is thinking}}$

Hints:

What is a suitable L?

What is L^*?

$\hat{u} = L^*(LL^*)^{-1}4 =?$

Properties of the Adjoint

Let L be a bounded linear operator between two real Hilbert spaces. Then (where the bar denotes 'closure')

$$
\begin{aligned}
L^{**} &= L && (1) \\
[\mathcal{R}(L)]^{\perp} &= \mathcal{N}(L^*) && (2) \\
[\mathcal{R}(L^*)]^{\perp} &= \mathcal{N}(L) && (3) \\
\overline{\mathcal{R}(L)} &= [\mathcal{N}(L^*)]^{\perp} && (4) \\
\overline{\mathcal{R}(L^*)} &= [\mathcal{N}(L)]^{\perp} && (5) \\
\mathcal{N}(L^*) &= \mathcal{N}(LL^*) && (6) \\
\mathcal{N}(L) &= \mathcal{N}(L^*L) && (7)
\end{aligned}
$$

Properties of the Adjoint

Proof of (2):

$$y \in \mathcal{R}(L)^{\perp} \Leftrightarrow < y, Lx > = 0, \ \forall x \Leftrightarrow < L^* y, x > = 0, \ \forall x \Leftrightarrow \ y \in N(L^*)$$

Proof of (7):

$$Lx = 0 \ \Rightarrow \ L^* Lx = 0 \ \Rightarrow \ 0 = < x, L^* Lx > = < Lx, Lx > \Rightarrow \ Lx = 0$$

Example: Shift Operator on l_2

$$
\begin{aligned}
l_2 &= \{x = (x_1, x_2, x_3, \ldots) : \ \textstyle\sum_{i=1}^{\infty} {x_i}^2 < \infty\} \\
x &= (x_1, x_2, x_3, \ldots) \\
y &= (y_1, y_2, y_3, \ldots) \\
Sx &= (0, x_1, x_2, \ldots) \\
S^* y &= (y_2, y_3, \ldots) \\
< y, Sx > &= \textstyle\sum_{i=1}^{\infty} y_{i+1} x_i =< S^* y, x > \\
\mathcal{R}(S) &= \{(0, *, *, *, \ldots)\} \\
\mathcal{N}(S^*) &= \{(*, 0, 0, 0, \ldots)\}
\end{aligned}
$$

Operator Interpretation of Gramian

Recall that the matrix function

$$W(t_0, t_f) = \int_{t_0}^{t_f} \Phi(t_0, t)B(t)B(t)^T\Phi(t_0, t)^T dt$$

is called *controllability Gramian*.

Define $L : \mathbf{L}_2^m[t_0, t_f] \to \mathbf{R}^n$ by $Lu = \int_{t_0}^{t_f} \Phi(t_0, \tau)B(\tau)u(\tau)d\tau$. Then

$$\begin{aligned} x(t_f) &= \Phi(t_f, t_0)[x(t_0) + Lu] \\ (L^*x)(t) &= B(t)^T\Phi(t_0, t)^T x \\ LL^* &= \int_{t_0}^{t_f} \Phi(t_0, \tau)B(\tau)B(\tau)^T\Phi(t_0, \tau)^T d\tau = W(t_0, t_f) \end{aligned}$$

Th. Rugh 9.2 - Controllability Revisited

The system $\dot{x}(t) = A(t)x(t) + B(t)u(t)$ is controllable on (t_0, t_f) if and only if $W(t_0, t_f) > 0$. The minimal cost $\int_{t_0}^{t_f} |u|^2 dt$ to reach 0 from x_0 is then $x_0^T W(t_0, t_f)^{-1} x_0$.

Proof.

$$
\begin{aligned}
\text{Reachability on } (t_0, t_f) &\Leftrightarrow && \forall x_0 : \exists u : x(t_f) = 0 \\
&\Leftrightarrow && \forall x_0 : \exists u : x_0 + Lu = 0 \\
&\Leftrightarrow && \mathcal{R}(L) = \mathbf{R}^n \\
&\Leftrightarrow && \mathcal{N}(L^*) = \{0\} \\
&\Leftrightarrow && \mathcal{N}(LL^*) = \{0\} \\
&\Leftrightarrow && \mathcal{N}[W(t_0, t_f)] = \{0\} \\
&\Leftrightarrow && W(t_0, t_f) > 0
\end{aligned}
$$

Controllability Cont'd

Minimize $|u|$ under the constraint $x_0 + Lu = 0$.

$$\begin{aligned} \hat{u} &= -L^*(LL^*)^{-1}x_0 \quad \text{(if } LL^* \text{ invertible)} \\ |\hat{u}|^2 &= x_0^T(LL^*)^{-1}x_0 = x_0^T W(t_0, t_f)^{-1}x_0 \end{aligned}$$

Observability Gramian

For $x_0 \in \mathbf{R}^n$, $y \in \mathbf{L}_2^m[t_0, t_1]$, introduce

$$\begin{aligned} (Mx_0)(t) &= C(t)\Phi(t,t_0)x_0, \quad t \in [t_0, t_1] \\ M^* y &= \int_{t_0}^{t_1} \Phi(t,t_0)^T C(t)^T y(t) dt \end{aligned}$$

Then the unobservable initial states can be computed as

$$\mathcal{N}(M) \;=\; \mathcal{N}(M^*M) = \mathcal{N}\left(\int_{t_0}^{t_1} \Phi(t,t_0)^T C(t)^T C(t) \Phi(t,t_0) dt\right)$$

Note that the matrix

$$\int_{t_0}^{t_1} \Phi(t,t_0)^T C(t)^T C(t) \Phi(t,t_0) dt$$

is the observability Gramian of the system.

Example - Polynomial Interpolation

Given m points (x_i, y_i) find a degree n polynomial

$$p(x) = p_0 + p_1 x + \dots p_n x^n$$

minimizing the interpolation error

$$J = \sum_{i=1}^{m} |y_i - p(x_i)|^2$$

Note that

$$\begin{pmatrix} p(x_1) \\ p(x_2) \\ \vdots \\ p(x_m) \end{pmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots x_1^n \\ 1 & x_2 & x_2^2 & \dots x_2^n \\ \vdots & & & \\ 1 & x_m & x_m^2 & \dots x_m^n \end{bmatrix} \begin{pmatrix} p_0 \\ p_1 \\ \vdots \\ p_n \end{pmatrix} := Lp$$

Example - Polynomial Interpolation

The problem is hence of the form: Find p that minimizes

$$|y - Lp|^2$$

The solution is given by (OP1)

$$\hat{p} = (L^* L)^{-1} L^* y$$
$$p(t) = \begin{pmatrix} 1 & t & \dots & t^n \end{pmatrix} \hat{p}$$

Example - Function Approximation

Given a set of basis functions $\Psi_i(x)$ and a function $v(x)$ solve the approximation problem

$$\min \int_a^b |v(x) - \sum_{i=1}^{n} u_i \Psi_i(x)|^2 dx$$

Solution $(L^* L)u = L^* v$ gives (check)

$$\begin{pmatrix} < \Psi_1, \Psi_1 > & \dots & < \Psi_1, \Psi_n > \\ \vdots & & \\ < \Psi_n, \Psi_1 > & \dots & < \Psi_n, \Psi_n > \end{pmatrix} u = \begin{pmatrix} < \Psi_1, v > \\ \vdots \\ < \Psi_n, v > \end{pmatrix}$$

Example - Function Approximation

Find a 2nd order polynomial approximating e^t for $0 \leq t \leq 1$

$$\min \int_0^1 |e^t - u_0 - u_1 t - u_2 t^2|^2 dt$$

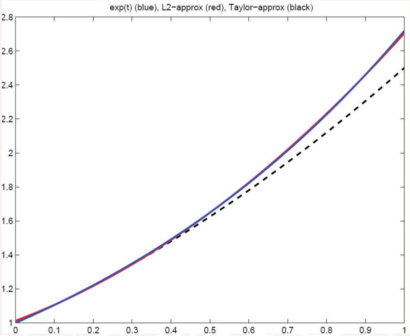
Calculation of $< t^k, t^m > = 1/(k+m+1)$ and $< t^k, e^t >$ gives

$$\begin{bmatrix} u_0 \\ u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}^{-1} \begin{bmatrix} e-1 \\ 1 \\ e-2 \end{bmatrix}$$

Giving the approximation

$$u(t) \approx 1.013 + 0.851t + 0.839t^2$$

Example - Function Approximation



Note that the L_2 approximation (red)

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

is significantly better than the Taylor approximation (black)

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