Engr210a Lecture 10: Hankel Operators and Model Reduction

- Hankel Operators
- Kronecker's theorem
- Discrete-time systems
- The Hankel norm
- Fundamental limitations
- Balanced realizations
- Balanced truncation

Hankel Operators

Suppose G has a minimal state-space system with $D = 0$. The operator

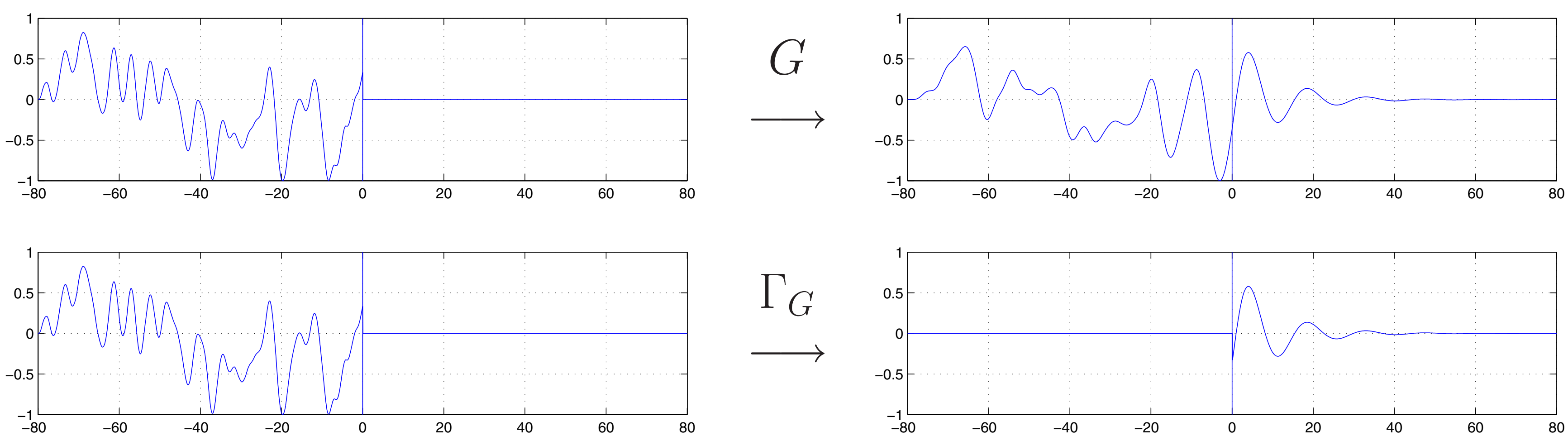
$$\Gamma_G : L_2(-\infty, 0] \to L_2[0, \infty) \qquad \text{defined by} \qquad \Gamma_G = P_+ G\big|_{L_2(-\infty,0]}$$

is called the *Hankel operator* corresponding to G.

- $P_+ : L_2(-\infty, \infty) \to L_2(-\infty, 0]$ is the projection operator

$$(P_+u)(t) = \begin{cases} 0 & \text{for } t < 0 \\ u(t) & \text{for } t \geq 0 \end{cases}$$

- $G\big|_{L_2(-\infty,0]}$ is G restricted to $L_2(-\infty, 0]$.



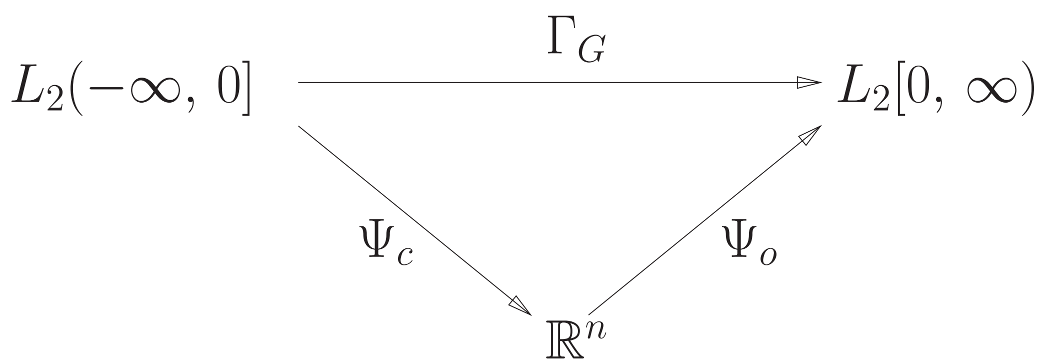
Hankel Operators

Suppose G has a minimal state-space system with $D = 0$. The operator

$$\Gamma_G : L_2(-\infty, 0] \to L_2[0, \infty) \qquad \text{defined by} \qquad \Gamma_G = P_+ G\big|_{L_2(-\infty,0]}$$

is called the *Hankel operator* corresponding to G.

Interpretation



- $\Gamma_G = \Psi_o \Psi_c$
- $\mathrm{rank}(\Gamma_G) \leq n$ for a state-space system of order n.
- Interpretation: the state summarizes all the information about the past inputs necessary to generate future outputs.

Operator rank

Suppose $A : \mathcal{U} \to \mathcal{V}$ is a map between Hilbert spaces $\mathcal{U}$ and $\mathcal{V}$. The *rank* of an operator A is defined by

$$\operatorname{rank}(A) = \dim\big(\operatorname{image}(A)\big)$$

Notes

If A has finite rank, then the following hold:

- $\operatorname{rank}(A) = \operatorname{rank}(A^*)$
- If $A : \mathbb{R}^n \to \mathcal{U}$, then $\operatorname{rank}(A) = n - \dim(\ker(A))$.
- $\operatorname{rank}(AB) = \operatorname{rank}(A^*AB)$. In particular, $\operatorname{rank}(A) = \operatorname{rank}(A^*A)$.

Controllability and Observability

- $\operatorname{rank}(Y_o) = \operatorname{rank}(\Psi_o^*\Psi_o) = \operatorname{rank}(\Psi_o) = n - \dim(\ker(\Psi_o))$
  $=$ dimension of the observable subspace
- $\operatorname{rank}(X_c) = \operatorname{rank}(\Psi_c\Psi_c^*) = \operatorname{rank}(\Psi_c) = \dim\big(\operatorname{image}(\Psi_c)\big)$
  $=$ dimension of the controllable subspace

Kronecker's theorem

Suppose G is a linear system with Hankel operator Γ_G, and suppose $\operatorname{rank}(\Gamma_G)$ is finite. Then a minimal realization of G has state-dimension equal to $\operatorname{rank}(\Gamma_G)$. Equivalently, for $A \in \mathbb{R}^{n\times n}$,

$$(A, B, C, D) \text{ is minimal} \qquad \iff \qquad \operatorname{rank}(\Gamma_G) = n$$

Proof

We will use the fact that
$$\begin{aligned}\operatorname{rank}(\Gamma_G) &= \operatorname{rank}(\Psi_o\Psi_c) = \operatorname{rank}(\Psi_o^*\Psi_o\Psi_c\Psi_c^*)\\ &= \operatorname{rank}(Y_oX_c)\end{aligned}$$

$\Longleftarrow$: Sylvester's inequality gives

$$\operatorname{rank}(\Gamma_G) = \operatorname{rank}(Y_oX_c) \le \min\{\operatorname{rank}(Y_o), \operatorname{rank}(X_c)\}$$

hence the system is controllable and observable

$\Longrightarrow$: The other Sylvester inequality gives

$$\begin{aligned}\operatorname{rank}(\Gamma_G) = \operatorname{rank}(Y_oX_c) &\ge \operatorname{rank}(Y_o) + \operatorname{rank}(X_c) - n\\ &= n\end{aligned}$$

Discrete-time systems

Suppose we have the state-space system

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) \end{aligned}$$

If G is stable, this defines a bounded linear operator $G : \ell_2(\mathbb{Z}_+) \to \ell_2(\mathbb{Z}_+)$. We can write an *infinite matrix* description for it as follows.

$$\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ y(3) \\ \vdots \end{bmatrix} = \begin{bmatrix} 0 & & & & \\ CB & 0 & & & \\ CAB & CB & 0 & & \\ CA^2B & CAB & CB & 0 & \\ \vdots & & & & \ddots \end{bmatrix} \begin{bmatrix} u(0) \\ u(1) \\ u(2) \\ u(3) \\ \vdots \end{bmatrix}$$

Notes

- The matrix G is structured; it is constant on diagonal from top-left to bottom-right. Such matrices are called *Toeplitz* matrices.
- G is Toeplitz if and only if G is time-invariant.
- G is lower-triangular if and only if G is causal.
- G is unchanged by changes in state-space coordinates.

Hankel operators in discrete-time

The controllability operator $\Psi_c : \ell_2(\mathbb{Z}_-) \to \mathbb{R}^n$ is given by

$$\xi = \Psi_c u \qquad \iff \qquad \xi = \begin{bmatrix} B & AB & A^2 & A^3B & \ldots \end{bmatrix} \begin{bmatrix} u(-1) \\ u(-2) \\ u(-3) \\ \vdots \end{bmatrix}$$

The observability operator $\Psi_o : \mathbb{R}^n \to \ell_2(\mathbb{Z}+)$ is given by

$$y = \Psi_o \xi \qquad \iff \qquad \begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ y(3) \\ \vdots \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ CA^3 \\ \vdots \end{bmatrix} \xi$$

Then the Hankel operator is

$$\Gamma_G = \Psi_o \Psi_c = \begin{bmatrix} CB & CAB & CA^2B & \ldots \\ CAB & CA^2B & CA^3B & \\ CA^2B & CA^3B & CA^4B & \\ CA^3B & CA^4B & CA^5B & \\ \vdots & & & \ddots \end{bmatrix}$$

Hankel operators in discrete-time

In discrete time, the Hankel operator is

$$\Gamma_G = \Psi_o \Psi_c = \begin{bmatrix} CB & CAB & CA^2B & \dots \\ CAB & CA^2B & CA^3B & \\ CA^2B & CA^3B & CA^4B & \\ CA^3B & CA^4B & CA^5B & \\ \vdots & & & \ddots \end{bmatrix}$$

Notes

- The infinite matrix Γ_G corresponding to the Hankel operator is constant along diagonals from top-right to bottom-left. Such a matrix is called a *Hankel matrix*.
- The coefficients along any row or column are the impulse response coefficients.
- Hence we can construct Γ_G from experimental data. This leads to a method of identification called *subspace identification*.
- Γ_G is unchanged by changes in state-space coordinates. Recall

$$C \to CT^{-1} \qquad A \to TAT^{-1} \qquad B \to TB$$

Hankel operators

- The Hankel Operator is $\Gamma_G = \Psi_o \Psi_c$, where

$$x = \Psi_c u \quad \Longrightarrow \quad x = \int_{-\infty}^{0} e^{-A\tau} B u(\tau)\, d\tau$$

$$y = \Psi_o x \quad \Longrightarrow \quad y(t) = Ce^{At} x$$

- Then we have

$$\begin{aligned} \Gamma_G u &= \int_{-\infty}^{0} Ce^{At-\tau} B u(\tau)\, d\tau \\ &= \int_{0}^{\infty} Ce^{A(t+\tau)} B u(-\tau)\, d\tau \end{aligned}$$

- In general, if G has impulse response g, then

$$u \in L_2[0,\infty),\ y = Gu \quad \Longrightarrow \quad y(t) = \int_0^t h(t-\tau) u(\tau)\, d\tau$$

$$u \in L_2(-\infty,0],\ y = \Gamma_G u \quad \Longrightarrow \quad y(t) = \int_0^\infty h(t+\tau) u(-\tau)\, d\tau$$

An integral operator with this structure is said to have *Hankel structure*.

The Hankel norm

The *Hankel norm* of the system G is the induced-norm of its Hankel operator. It satisfies

$$\|\Gamma_G\| = \big(\lambda_{\mathsf{max}}(Y_o X_c)\big)^{\frac{1}{2}}$$

In fact $\operatorname{spec}(\Gamma_G^* \Gamma_G) = \operatorname{spec}(Y_o X_c) \cup \{0\}$.

Proof

- We know $\|\Gamma_G\| = \|\Gamma_G^* \Gamma_G\|^{\frac{1}{2}} = \big(\rho(\Gamma_G^* \Gamma_G)\big)^{\frac{1}{2}}$
- Also $$\begin{aligned} \operatorname{spec}(\Gamma_G^* \Gamma_G) &= \operatorname{spec}(\Psi_c^* \Psi_o^* \Psi_o \Psi_c) \\ &= \operatorname{spec}(\Psi_o^* \Psi_o \Psi_c \Psi_c^*) \cup \{0\} \\ &= \operatorname{spec}(Y_o X_c) \cup \{0\} \end{aligned}$$
- The eigenvalues of $Y_o X_c$ are real and positive, since $\operatorname{spec}(Y_o X_c) = \operatorname{spec}(X_c^{\frac{1}{2}} Y_o X_c^{\frac{1}{2}})$.

Notes

- The square-roots of the eigenvalues of $\Gamma_G^* \Gamma_G$ are called the *Hankel singular values* of G. They are usually written $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$. Zero is not included.
- The Hankel singular values are independent of the state-space coordinates.

Coordinate invariance

- The controllability and observability gramians depend on the choice of coordinates in state-space.
- However, the Hankel singular values are independent of the state-space coordinates.
- If $z = Tx$, then (A, B, C, D) transforms to $(\tilde{A}, \tilde{B}, \tilde{C}, D)$ where $\tilde{A} = TAT^{-1}$, $\tilde{B} = TB$, $\tilde{C} = CT^{-1}$.
- $x = \Psi_c u$ implies $z = T\Psi_c u$, hence $\tilde{\Psi}_c = T\Psi_c$. Hence

$$\begin{aligned} \tilde{X}_c &= \tilde{\Psi}_c \tilde{\Psi}_c^* = T\Psi_c\Psi_c^* T^* \\ &= TX_cT^* \end{aligned}$$

  Similarly, $\tilde{\Psi}_o = \Psi_o T^{-1}$ implies $\tilde{Y}_o = (T^*)^{-1} Y_o T^{-1}$.

- As expected, $\tilde{\Gamma}_G = \tilde{\Psi}_o \tilde{\Psi}_c = \Psi_o T^{-1} T \Psi_c = \Psi_o \Psi_c = \Gamma_G$.
- Also

$$\begin{aligned} \operatorname{spec}(\tilde{Y}_o \tilde{X}_c) &= \operatorname{spec}\big((T^*)^{-1} Y_o T^{-1} T X_c T^*\big) \\ &= \operatorname{spec}\big((T^*)^{-1} Y_o X_c T^*\big) \\ &= \operatorname{spec}(Y_o X_c) \end{aligned}$$

Hankel Norm

The Hankel norm satisfies

$$\|\Gamma_G\| \leq \|G\|$$

Proof

The projection P_+ has norm $\|P_+\| = 1$. Hence

$$\begin{aligned}\|\Gamma_G\| &= \left\| P_+ G \big|_{L_2(-\infty,0]} \right\| \\ &\leq \|P_+\| \left\| G \big|_{L_2(-\infty,0]} \right\| \\ &= \left\| G \big|_{L_2(-\infty,0]} \right\| \\ &\leq \|G\|\end{aligned}$$

Interpretation

- $\|G\| = \sup_{\|u\|=1} \|Gu\|$, the maximum norm of the total output
- $\|\Gamma_G\| = \sup_{\|u\|=1} \|\Gamma_G u\|$, the maximum norm of the output on $t > 0$.

Model reduction

Suppose $G \in H_\infty$ has a minimal realization of dimension n. Given $r < n$, we would like to find the $G_r \in H_\infty$ which solves

$$\begin{array}{rl} \text{minimize} & \|G - G_r\| \\ \text{subject to} & G_r \text{ has state-dimension } r \end{array}$$

Notes

- For any G and G_r,

  $$\|G - G_r\| \geq \|\Gamma_{G-G_r}\| = \|\Gamma_G - \Gamma_{G_r}\|$$

  This leads to the problem of optimal Hankel norm approximation

Optimal Hankel-norm approximation

Given Γ_G, find an operator $\Gamma_{G_r} : L_2(-\infty, 0] \to L_2[0, -\infty)$ which solves

$$\begin{array}{rl} \text{minimize} & \|\Gamma_G - \Gamma_{G_r}\| \\ \text{subject to} & \Gamma_{G_r} \text{ is the Hankel operator for some } G_r \in H_\infty \\ & \operatorname{rank}(\Gamma_{G_r}) = r \end{array}$$

Optimal Hankel-norm approximation

Given Γ_G, find an operator $\Gamma_{G_r} : L_2(-\infty, 0] \to L_2[0, -\infty)$ which solves

$$\begin{aligned} \text{minimize} \quad & \|\Gamma_G - \Gamma_{G_r}\| \\ \text{subject to} \quad & \Gamma_{G_r} \text{ is the Hankel operator for some } G_r \in H_\infty \\ & \operatorname{rank}(\Gamma_{G_r}) = r \end{aligned}$$

Notes

- Suppose $\Gamma_{G_{\text{hankel-optimal}}}$ is the optimal. Then for any system G_r of order r,

$$\begin{aligned} \|G - G_r\| &\geq \|\Gamma_G - \Gamma_{G_r}\| \\ &\geq \|\Gamma_G - \Gamma_{G_{\text{hankel-optimal}}}\| \end{aligned}$$

- So if we can solve the optimal Hankel-norm approximation problem, then we have a lower-bound on the best-possible error achievable in the induced-norm for the model reduction problem.

Optimal Hankel-norm approximation

Given Γ_G, find an operator $\Gamma_{G_r} : L_2(-\infty, 0] \to L_2[0, -\infty)$ which solves

$$\begin{aligned} \text{minimize} \quad & \|\Gamma_G - \Gamma_{G_r}\| \\ \text{subject to} \quad & \Gamma_{G_r} \text{ is the Hankel operator for some } G_r \in H_\infty \\ & \operatorname{rank}(\Gamma_{G_r}) = r \end{aligned}$$

Relaxed problem

Given Γ_G, find an operator $W : L_2(-\infty, 0] \to L_2[0, -\infty)$ which solves

$$\begin{aligned} \text{minimize} \quad & \|\Gamma_G - W\| \\ \text{subject to} \quad & \operatorname{rank}(W) = r \end{aligned}$$

Notes

- This is just a minimal-rank approximation problem; for matrices we can use SVD.
- We have $\|\Gamma_G - \Gamma_{G_{\text{hankel-optimal}}}\| \geq \|\Gamma_G - W_{\text{opt}}\|$, since in general W_{opt} will not have Hankel structure.
- Hence for any system G_r of order r,

$$\|G - G_r\| \geq \|\Gamma_G - W_{\text{opt}}\|$$

Minimal rank r matrix approximation

Recall the optimal rank approximation problem. Given $A \in \mathbb{C}^{m \times n}$,

$$\begin{array}{ll} \text{minimize} & \|A - B\| \\ \text{subject to} & \operatorname{rank}(B) = r \end{array}$$

Singular value decomposition

Given $A \in \mathbb{C}^{m \times n}$, we can decompose it into the *singular value decomposition* (SVD)

$$A = U \Sigma V^*$$

where $U \in \mathbb{C}^{m \times m}$ is unitary, $V \in \mathbb{C}^{n \times n}$ is unitary, $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal.

Notes

- $\Sigma_{ii} = \sigma_i$, ordered so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{m,n\}}$.
- The optimal B satisfies $\|A - B_{\text{opt}}\| = \sigma_{k+1}$.
- $B_{\text{opt}} = \displaystyle\sum_{i=1}^{k} \sigma_i u_i v_i^*$

Theorem

Suppose G has a minimal realization of order n. Then for any G_r of order $r < n$,

$$\|G - G_r\| \geq \sigma_{r+1}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$ are the Hankel singular values of G.

Proof

- We show that $\|\Gamma_G - W\| \geq \sigma_{r+1}$ if $\mathrm{rank}(W) = r$.
- Let $\Gamma_G = \Psi_o \Psi_c$, and define $P_o : L_2[0, \infty) \to \mathbb{C}^n$ and $P_c : \mathbb{C}^n \to L_2(-\infty, 0]$ by

  $$P_o = Y_o^{-\frac{1}{2}} \Psi_o^* \qquad P_c = \Psi_c^* X_c^{-\frac{1}{2}}$$

  Note that $\|P_o\| = \|P_c\| = 1$. Then

  $$\begin{aligned}\|\Gamma_G - W\| &= \|P_o\| \|\Gamma_G - W\| \|P_c\| \geq \|P_o(\Gamma_G - W)P_c\| \\ &= \|Y_o^{-\frac{1}{2}} \Psi_o^* \Psi_o \Psi_c \Psi_c^* X_c^{-\frac{1}{2}} - P_o W P_c\| \\ &= \|Y_o^{\frac{1}{2}} X_c^{\frac{1}{2}} - P_o W P_c\|\end{aligned}$$

- $\mathrm{rank}(P_o W P_c) \leq r$, since $\mathrm{rank}(W) \leq r$, hence

  $$\|Y_o^{\frac{1}{2}} X_c^{\frac{1}{2}} - P_o W P_c\| \geq \sigma_{r+1}(Y_o^{\frac{1}{2}} X_c^{\frac{1}{2}}) = \left(\lambda_{r+1}(Y_o^{\frac{1}{2}} X_c Y_o^{\frac{1}{2}})\right)^{\frac{1}{2}} = \sigma_{r+1}$$

Bounds on the model reduction error

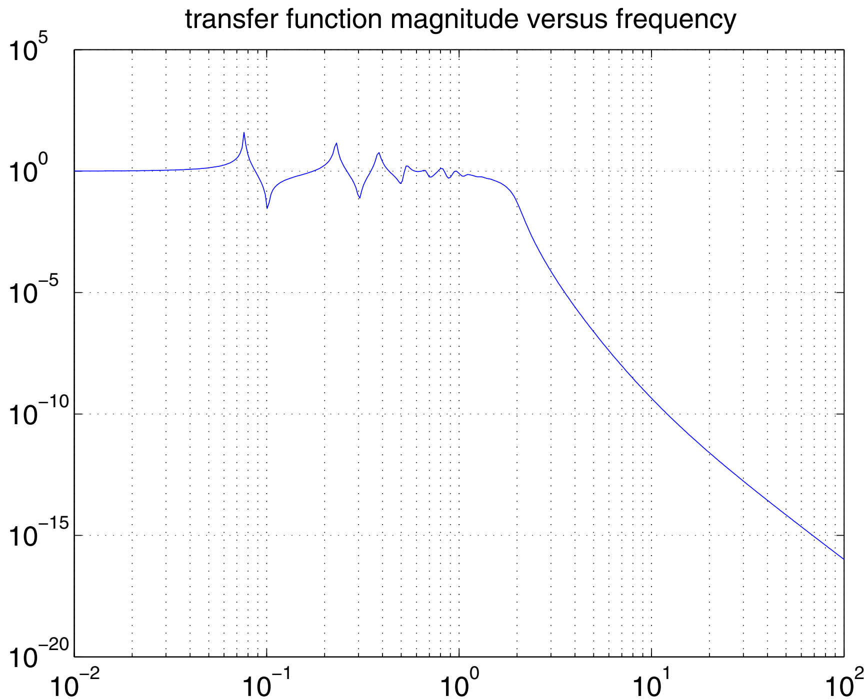
We have seen the lower-bound

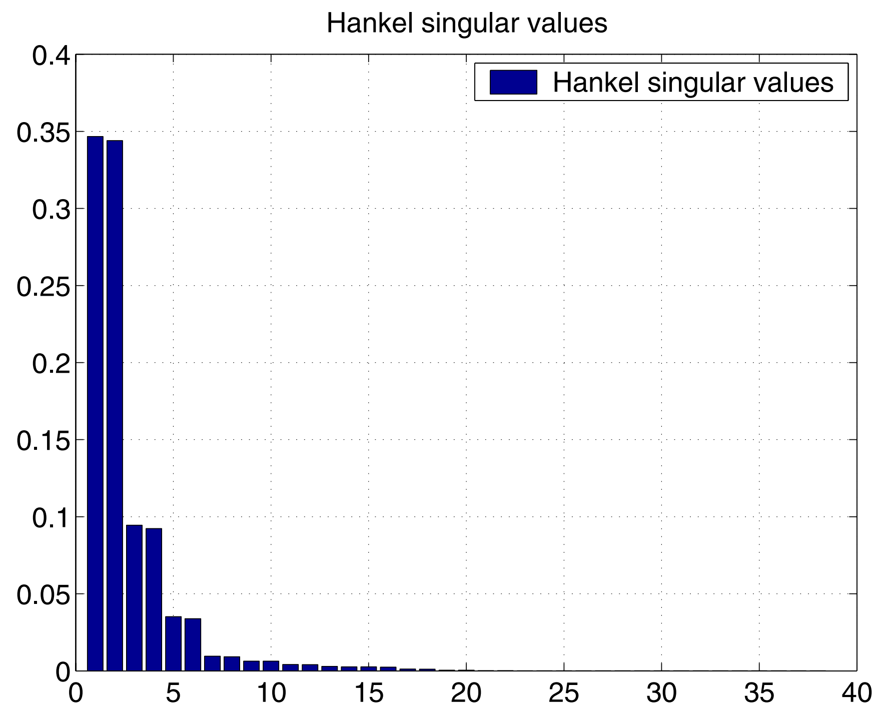
$$\|G - G_r\| \geq \sigma_{r+1}$$

No G_r of order r can do better than this.

Example

Mechanical system with state-dimension 40.





Ellipsoids example

$$A = \begin{bmatrix} 0 & -1.25 \\ 4 & -6 \end{bmatrix} \qquad B = \begin{bmatrix} 1.2 \\ 4.48 \end{bmatrix}$$

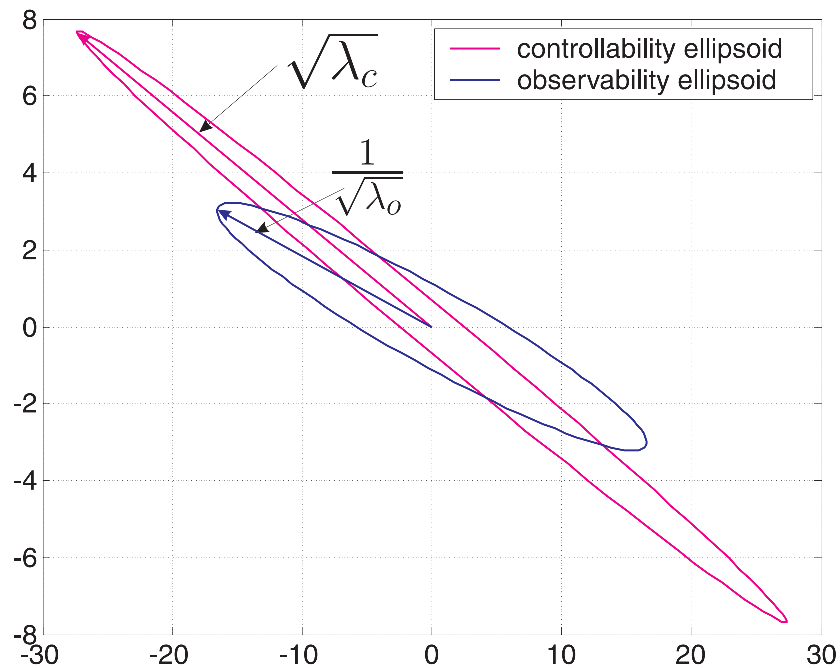
$$C = \begin{bmatrix} 20 & 0 \end{bmatrix}$$

The controllability and observability ellipsoids are

$$E_c = \left\{ \Psi_c u \ ; \ \|u\| \le 1 \right\} = \left\{ x \in \mathbb{R}^n \ ; \ x^* X_c^{-1} x \le 1 \right\}$$
$$E_o = \left\{ x \in \mathbb{R}^n \ ; \ \|\Psi_o x\| \le 1 \right\} = \left\{ x \in \mathbb{R}^n \mid x^* Y_o x \le 1 \right\}$$

Notes

- The ellipsoids are almost aligned.
- Even though some states are weakly observable, they are also strongly controllable.
- Input-to-state map Ψ_c has worst-case scaling $\sqrt{\lambda_c}$.
- State-to-output map Ψ_o has worst-case scaling $\sqrt{\lambda_o}$.



Balanced realizations

Recall that under state-transformation T,

$$X_c \to T X_c T^* \qquad Y_o \to (T^*)^{-1} Y_o T^{-1}$$

If the realization (A, B, C, D) is controllable and observable, then we can choose state-space coordinates in which the controllability and observability gramians are equal and diagonal. A realization with this property is called a *balanced realization.*

Construction

- Using the eigenvalue decomposition for symmetric matrices (or SVD)

  $$X^{\frac{1}{2}} Y X^{\frac{1}{2}} = U \Sigma^2 U^*$$

  where U is unitary and Σ is diagonal, positive definite.

- Hence $\Sigma^{-\frac{1}{2}} U^* X^{\frac{1}{2}} Y X^{\frac{1}{2}} U \Sigma^{-\frac{1}{2}} = \Sigma$
- Let $T^{-1} = X^{\frac{1}{2}} U \Sigma^{-\frac{1}{2}}$. Then the above states that $(T^{-1})^* Y T^{-1} = \Sigma$. Also

  $$T X T^* = (\Sigma^{\frac{1}{2}} U^* X^{-\frac{1}{2}})\, X\, (X^{-\frac{1}{2}} U \Sigma^{\frac{1}{2}}) = \Sigma.$$

- Hence in the new coordinates, $X_c = Y_o = \Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$.

Balanced realizations

- Every system $G \in H_\infty$ has a minimal balanced realization.
- In the balanced realization, the controllability and observability Gramians are equal. Hence strongly controllable states are also strongly observable, and weakly controllable states are also weakly observable.

$$X_c = Y_o = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}$$

- We can always choose the ordering so that $\sigma_i \geq \sigma_{i+1}$.
- Hence we might expect that removing the weakly observable and weakly controllable states would result in a low model-reduction error. This turns out to be the case.

Balanced truncation

Given G or order n, we wish to find a reduced-order model of order $r < n$. Suppose $D = 0$, and A, B, C is a balanced realization for G. Partition matrices A, B, C as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \qquad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \quad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}$$

where $A_{11} \in \mathbb{R}^{r \times r}$. The reduced order model will be

$$\hat{G}_r(s) = \left[\begin{array}{c|c} A_{11} & B_1 \\ \hline C_1 & 0 \end{array}\right]$$

This reduced-order model is called a *balanced truncation* of G.

Notes

- Assume $\sigma_r > \sigma_{r+1}$. That is, these singular values must not be equal.
- We will show that G_r is stable and balanced, and derive an upper bound on the modeling error

$$\|G - G_r\|$$

- The method of *truncation* is an example of a *Galerkin projection* of the differential equations onto a particular basis; the basis we are using is that spanned by the r most controllable and observable states.