Selected Topics in Numerical Linear Algebra and Control

Distance Concepts

Zlatko Drmač and Daniel Kressner
{drmac,kressner}math.hr

Recall notions of stability/controllability/observability for LTI continuous time control system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \qquad x(0) = x_0, \\ y(t) &= Cx(t) + Du(t), \end{aligned}$$

where

$A \in \mathbb{R}^{n \times n}$ = state matrix,
$B \in \mathbb{R}^{n \times m}$ = input matrix,
$C \in \mathbb{R}^{p \times m}$ = output matrix,
$D \in \mathbb{R}^{p \times m}$ = feedthrough matrix.

Asymptotic stability	$\Leftrightarrow$	$\lambda(A) \subset \mathbb{C}^-$
Controllability	$\Leftrightarrow$	$\text{rank}([B, AB, A^2B, \cdots, A^{n-1}B]) = n$
Observabilty	$\Leftrightarrow$	$\text{rank}([C; CA; CA^2; \cdots; CA^{n-1}]) = n$

These notions make sense when the system is given exactly.
But *virtually always*, the data describing the system is corrupted, due to:

- modelling simplifications;
- discretization and linearization;
- other approximation errors;
- neglected stochastic parts (e.g., white noise);
- roundoff error.

$$\begin{aligned} \text{Asymptotic stability} \quad &\Leftrightarrow \quad \lambda(A) \subset \mathbb{C}^- \\ \text{Controllability} \quad &\Leftrightarrow \quad \text{rank}([B, AB, A^2B, \cdots, A^{n-1}B]) = n \\ \text{Observabilty} \quad &\Leftrightarrow \quad \text{rank}([C; CA; CA^2; \cdots; CA^{n-1}]) = n \end{aligned}$$

These notions make sense when the system is given exactly.
But *virtually always*, the data describing the system is corrupted, due to:

- modelling simplifications;
- discretization and linearization;
- other approximation errors;
- neglected stochastic parts (e.g., white noise);
- roundoff error.

It is more appropriate to ask whether a given property holds for all systems in a neighborhood of the given system.

Tempting heuristics can be misleading!
(and often have little mathematical justification)

Demmel's matrix:

$$A = \begin{bmatrix} -1 & -5 & -25 & -125 & -625 \\ 0 & -1 & -5 & -25 & -125 \\ 0 & 0 & -1 & -5 & -25 \\ 0 & 0 & 0 & -1 & -5 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

All eigenvalues are pleasantly far away from the imaginary axis. But...

It is more appropriate to ask whether a given property holds for all systems in a neighborhood of the given system.

Tempting heuristics can be misleading!
(and often have little mathematical justification)

Demmel's matrix:

$$A = \begin{bmatrix} -1 & -5 & -25 & -125 & -625 \\ 0 & -1 & -5 & -25 & -125 \\ 0 & 0 & -1 & -5 & -25 \\ 0 & 0 & 0 & -1 & -5 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

All eigenvalues are pleasantly far away from the imaginary axis. But...

It is more appropriate to ask whether a given property holds for all systems in a neighborhood of the given system.

Tempting heuristics can be misleading!
(and often have little mathematical justification)

Demmel's matrix:

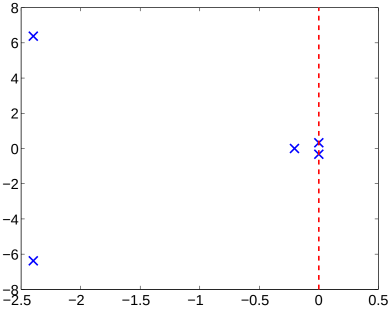
$$A = \begin{bmatrix} -1 & -5 & -25 & -125 & -625 \\ 0 & -1 & -5 & -25 & -125 \\ 0 & 0 & -1 & -5 & -25 \\ 0 & 0 & 0 & -1 & -5 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

All eigenvalues are pleasantly far away from the imaginary axis. But...

Slightly perturbed version of Demmel's matrix:

$$A+E=\begin{bmatrix} -1 & -5 & -25 & -125 & -625 \\ 0 & -1 & -5 & -25 & -125 \\ 0 & 0 & -1 & -5 & -25 \\ 0 & 0 & 0 & -1 & -5 \\ 0.06 & 0 & 0 & 0 & -1 \end{bmatrix}$$

$\lambda(A+E)$ (two eigenvalues have positive real part):



In the case of nonnormal matrices, eigenvalues do not provide a robust measure for stability.
More appropriate to consider distance to instability:

$$\beta(A) := \inf\{\|E\| : E \in \mathbb{C}^{n\times n}, A + E \text{ is not stable}\}$$

(A matrix is called *stable* if its eigenvalues are in the open left half plane.)

In the following, we assume that A is a stable matrix (otherwise, $\beta(A) = 0$.

In the case of nonnormal matrices, eigenvalues do not provide a robust measure for stability.
More appropriate to consider distance to instability:

$$\beta(A) := \inf\{\|E\| : E \in \mathbb{C}^{n\times n}, A + E \text{ is not stable}\}$$

(A matrix is called *stable* if its eigenvalues are in the open left half plane.)

In the following, we assume that A is a stable matrix (otherwise, $\beta(A) = 0$.

Continuity of eigenvalues implies that the smallest perturbation realizing $\beta(A)$ must move one of the eigenvalues to the imaginary axis. If $\|\cdot\|$ denotes the matrix 2-norm:

$$\begin{aligned}\beta(A) &= \inf\{\|E\| : E \in \mathbb{C}^{n\times n},\ \lambda(A+E)\cap \mathrm{i}\mathbb{R} \neq \emptyset\}\\ &= \inf\{\|E\| : E \in \mathbb{C}^{n\times n},\ \exists y \in \mathbb{R} : \det(A+E-\mathrm{i}yI)=0\}\\ &= \inf_{y\in\mathbb{R}} \inf\{\|E\| : E \in \mathbb{C}^{n\times n},\ \det(A+E-\mathrm{i}yI)=0\}\\ &= \inf_{y\in\mathbb{R}} \sigma_{\min}(A-\mathrm{i}yI)\end{aligned}$$

(Here, we used the fact that the norm of the smallest perturbation that makes a matrix B singular is $\sigma_{\min}(B)$, the smallest singular value of B.)

Continuity of eigenvalues implies that the smallest perturbation realizing $\beta(A)$ must move one of the eigenvalues to the imaginary axis. If $\|\cdot\|$ denotes the matrix 2-norm:

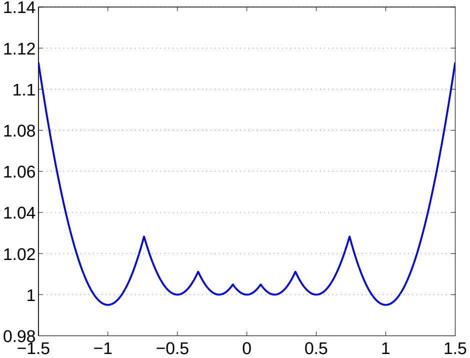
$$\begin{aligned}
\beta(A) &= \inf\{\|E\| : E \in \mathbb{C}^{n\times n},\ \lambda(A+E)\cap \mathrm{i}\mathbb{R} \neq \emptyset\} \\
&= \inf\{\|E\| : E \in \mathbb{C}^{n\times n},\ \exists y \in \mathbb{R} : \det(A+E-\mathrm{i}yI) = 0\} \\
&= \inf_{y\in\mathbb{R}} \inf\{\|E\| : E \in \mathbb{C}^{n\times n},\ \det(A+E-\mathrm{i}yI) = 0\} \\
&= \inf_{y\in\mathbb{R}} \sigma_{\min}(A-\mathrm{i}yI)
\end{aligned}$$

(Here, we used the fact that the norm of the smallest perturbation that makes a matrix B singular is $\sigma_{\min}(B)$, the smallest singular value of B.)

$$\beta(A) = \inf_{y \in \mathbb{R}} \sigma_{\min}(A - \mathrm{i}yI)$$

is one-parameter optimization problem. But not convex and only semi-smooth.

Graph of $\sigma_{\min}(A - \mathrm{i}yI)$ for 10×10 matrix with all eigenvalues having real part -1:



All global optimization methods for computing $\beta(A)$ are based on the following theorem.

Theorem

For real numbers x and y, $\epsilon > 0$ is a singular value of $A - (x + \mathrm{i}y)I$ if and only if $\mathrm{i}y$ is an eigenvalue of

$$H_x(\epsilon) = \begin{bmatrix} xI - A^* & \epsilon I \\ -\epsilon I & A - xI \end{bmatrix}.$$

Immediate consequences:

- If $H_0(\epsilon)$ has a purely imaginary eigenvalue $\mathrm{i}y$ then ϵ is a singular value of $A - \mathrm{i}yI$ $\rightsquigarrow$ $\epsilon \geq \beta(A)$.
- If $H_0(\epsilon)$ has no purely imaginary eigenvalue then ϵ is not a singular value of $A - \mathrm{i}yI$ for all y. But $\sigma_{\min}(A - \mathrm{i}yI)$ is continuous and $\rightarrow \pm\infty$ as $y \rightarrow \pm\infty$ $\rightsquigarrow$ $\epsilon < \beta(A)$

All global optimization methods for computing $\beta(A)$ are based on the following theorem.

Theorem

For real numbers x and y, $\epsilon > 0$ is a singular value of $A - (x + \mathrm{i}y)I$ if and only if $\mathrm{i}y$ is an eigenvalue of

$$H_x(\epsilon) = \begin{bmatrix} xI - A^* & \epsilon I \\ -\epsilon I & A - xI \end{bmatrix}.$$

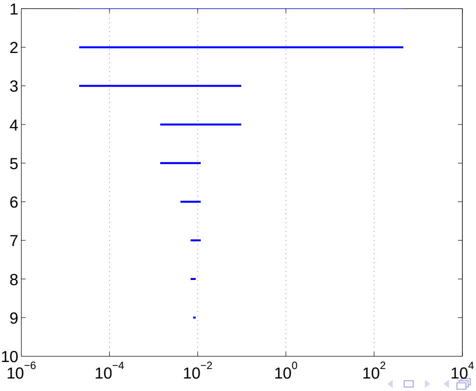
Immediate consequences:

- If $H_0(\epsilon)$ has a purely imaginary eigenvalue $\mathrm{i}y$ then ϵ is a singular value of $A - \mathrm{i}yI \rightsquigarrow \epsilon \geq \beta(A)$.
- If $H_0(\epsilon)$ has no purely imaginary eigenvalue then ϵ is not a singular value of $A - \mathrm{i}yI$ for all y. But $\sigma_{\min}(A - \mathrm{i}yI)$ is continuous and $\rightarrow \pm\infty$ as $y \rightarrow \pm\infty \rightsquigarrow \epsilon < \beta(A)$

Byers' bisection method

1. Set $l = 0$ and $u = \|A + A^*\|_F/2$.
2. Set $\epsilon = \sqrt{lu}$ and compute $\Lambda = \lambda(H_\epsilon)$.
3. If $\Lambda \cap \mathrm{i}\mathbb{R} = \emptyset$ then set $l = \epsilon$, otherwise set $u = \epsilon$.
4. If $|u - l|$ too large, go to Step 2.

For Demmel's matrix:



Convergence of bisection is too slow if high accuracy is needed (e.g., in algorithms for optimizing robust stability).

The Boyd-Balakrishnan method starts with an upper bound u and computes all imaginary eigenvalues $\mathrm{i}y_1, \ldots, \mathrm{i}y_m$ of $H_0(u)$ (including multiplicities) s.t. $y_1 \leq \cdots \leq y_m$.

Then the new u is obtained as

$$u = \min\ \sigma_{\min}(A - \mathrm{i}(y_{2j} - y_{2j-1})/2I), \quad j = 1, \ldots, m/2.$$

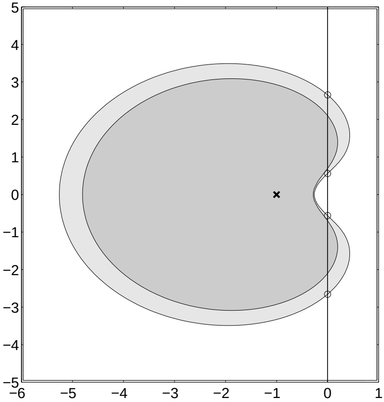
Convergence of bisection is too slow if high accuracy is needed (e.g., in algorithms for optimizing robust stability).

The Boyd-Balakrishnan method starts with an upper bound u and computes all imaginary eigenvalues $\mathrm{i}y_1, \ldots, \mathrm{i}y_m$ of $H_0(u)$ (including multiplicities) s.t. $y_1 \le \cdots \le y_m$.

Then the new u is obtained as

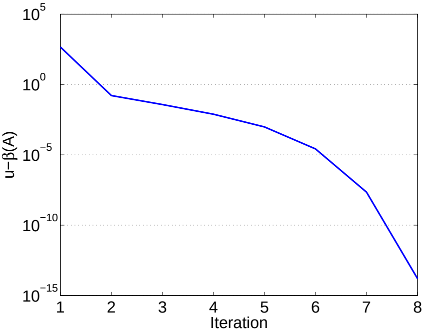
$$u = \min \ \sigma_{\min}(A - \mathrm{i}(y_{2j} - y_{2j-1})/2I), \quad j = 1, \ldots, m/2.$$

Easier to understand when looking at the pseudospectrum of A.
Consider pseudospectrum of Demmel's matrix for $\epsilon = 0.015$:



The imaginary axis crosses the boundary of the pseudospectrum (outer area) exactly at the points $\mathrm{i}y_1, \ldots, \mathrm{i}y_4$.

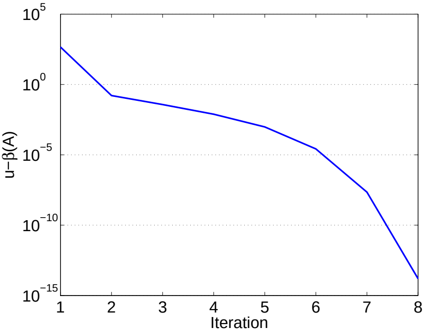
Convergence of Boyd-Balakrishnan method for Demmel's matrix:



In general:

- global convergence;
- quadratic local convergence.

Convergence of Boyd-Balakrishnan method for Demmel's matrix:



In general:

- global convergence;
- quadratic local convergence.

Computation of H_∞ norm

The H^∞ norm of an LTI continous-time control system with transfer matrix G was defined (for $D = 0$) as

$$\begin{aligned} \|G\|_{H^\infty} &= \sup_{\omega \in \mathbb{R}} \|G(\mathrm{i}\omega)\|_2 \\ &= \sup_{\omega \in \mathbb{R}} \|C(\mathrm{i}\omega I - A)^{-1} B\|_2. \end{aligned}$$

This already looks rather similar to $\beta(A) = \min_{y \in \mathbb{R}} \sigma_{\min}(A - \mathrm{i}yI)$!

In fact, it can be shown that $1/\gamma$ is a singular value of $C(\mathrm{i}\omega I - A)^{-1}B$ if and only if $\mathrm{i}\omega$ is an eigenvalue of

$$\begin{bmatrix} A^* & -\gamma BB^* \\ \gamma C^* C & -A \end{bmatrix}$$

$\rightsquigarrow$ Byers and Boyd-Balakrishnan methods can be extended.

Computation of H_∞ norm

The H^∞ norm of an LTI continous-time control system with transfer matrix G was defined (for $D = 0$) as

$$\begin{aligned} \|G\|_{H^\infty} &= \sup_{\omega\in\mathbb{R}} \|G(\mathrm{i}\omega)\|_2 \\ &= \sup_{\omega\in\mathbb{R}} \|C(\mathrm{i}\omega I - A)^{-1}B\|_2. \end{aligned}$$

This already looks rather similar to $\beta(A) = \min_{y\in\mathbb{R}} \sigma_{\min}(A - \mathrm{i}yI)$!

In fact, it can be shown that $1/\gamma$ is a singular value of $C(\mathrm{i}\omega I - A)^{-1}B$ if and only if $\mathrm{i}\omega$ is an eigenvalue of

$$\begin{bmatrix} A^* & -\gamma BB^* \\ \gamma C^*C & -A \end{bmatrix}$$

$\rightsquigarrow$ Byers and Boyd-Balakrishnan methods can be extended.

Computation of H_∞ norm

The H^∞ norm of an LTI continous-time control system with transfer matrix G was defined (for $D = 0$) as

$$\begin{aligned} \|G\|_{H^\infty} &= \sup_{\omega\in\mathbb{R}} \|G(\mathrm{i}\omega)\|_2 \\ &= \sup_{\omega\in\mathbb{R}} \|C(\mathrm{i}\omega I - A)^{-1}B\|_2. \end{aligned}$$

This already looks rather similar to $\beta(A) = \min_{y\in\mathbb{R}} \sigma_{\min}(A - \mathrm{i}yI)$!

In fact, it can be shown that $1/\gamma$ is a singular value of $C(\mathrm{i}\omega I - A)^{-1}B$ if and only if $\mathrm{i}\omega$ is an eigenvalue of

$$\begin{bmatrix} A^* & -\gamma BB^* \\ \gamma C^*C & -A \end{bmatrix}$$

$\rightsquigarrow$ Byers and Boyd-Balakrishnan methods can be extended.

Other extensions

In many situations, it is unrealistic to consider measures that are robust w.r.t. all perturbations in $\mathbb{C}^{n\times n}$. For example,

- if A is real then $A+E$ should stay real;
- if the system is an interconnected system, it makes much sense only to consider only perturbations in the subsystems (i.e., block diagonal perturbations).

These are, roughly speaking, the only two practically relevant cases that can be computationally handled.

μ-toolbox in MATLAB: $\sigma_{\min}(\cdot)$ is replaced by the μ-function.

Other extensions

In many situations, it is unrealistic to consider measures that are robust w.r.t. all perturbations in $\mathbb{C}^{n \times n}$. For example,

- if A is real then $A + E$ should stay real;
- if the system is an interconnected system, it makes much sense only to consider only perturbations in the subsystems (i.e., block diagonal perturbations).

These are, roughly speaking, the only two practically relevant cases that can be computationally handled.

μ-toolbox in MATLAB: $\sigma_{\min}(\cdot)$ is replaced by the μ-function.

Checking controllability

To check controllability, a good heuristic is often obtained by computing the controllability form. If system is controllable and $m = 1$ then $\exists$ orthogonal matrix s.t.

$$Q^T A Q = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \end{bmatrix}, \quad Q^T B = \begin{bmatrix} \times \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

If one of the $\times$'s is small, (A, B) is close to being uncontrollable.

Unfortunately, the converse...

Checking controllability

To check controllability, a good heuristic is often obtained by computing the controllability form. If system is controllable and $m = 1$ then $\exists$ orthogonal matrix s.t.

$$Q^T A Q = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \end{bmatrix}, \quad Q^T B = \begin{bmatrix} \times \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

If one of the $\times$'s is small, (A, B) is close to being uncontrollable.

Unfortunately, the converse...

is not true.

$$A = \begin{bmatrix} 6 & 5 & 4 & 3 & 2 & 1 \\ 0.01 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0.01 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0.01 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0.01 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0.01 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

but

$$\sigma_{\min}([A - 5I, B]) \approx 3 \times 10^{-12}.$$

is not true.

$$A = \begin{bmatrix} 6 & 5 & 4 & 3 & 2 & 1 \\ 0.01 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0.01 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0.01 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0.01 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0.01 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

but

$$\sigma_{\min}([A - 5I, B]) \approx 3 \times 10^{-12}.$$

More appropriate to consider distance to uncontrollability:

$$\delta(A, B) := \inf\{\|[E, F]\| : (A + E, B + F) \text{ is not controllable.}\}$$

From the controllabilty test

$$\text{rank}([A - \lambda I, B]) = n \qquad \forall \lambda \in \mathbb{C},$$

we obtain Eising's formula

$$\delta(A, B) = \min_{\lambda \in \mathbb{C}} \sigma_{\min}([A - \lambda I, B]).$$

This is a two-parameter optimization problem (much harder than distance to instability). Best known algorithm is $O(n^5)$ (in worst case).

More appropriate to consider distance to uncontrollability:

$$\delta(A, B) := \inf\{\|[E, F]\| : \ (A + E, B + F) \text{ is not controllable.}\}$$

From the controllabilty test

$$\text{rank}([A - \lambda I, B]) = n \qquad \forall \lambda \in \mathbb{C},$$

we obtain Eising's formula

$$\delta(A, B) = \min_{\lambda \in \mathbb{C}} \sigma_{\min}([A - \lambda I, B]).$$

This is a two-parameter optimization problem (much harder than distance to instability). Best known algorithm is $O(n^5)$ (in worst case).