Engr210a Lecture 3: Singular Values and LMIs

- Matrix norm
- Singular value decomposition (SVD)
- Minimal-rank approximation
- Sensitivity of eigenvalues and singular values
- Linear matrix inequalities (LMIs)
- Semidefinite programming problems

Norm of a matrix

Suppose $A \in \mathbb{R}^{m \times n}$. The *matrix norm* of A is

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

Also called the *operator norm* or *spectral norm*.

Gives the maximum *gain* or *amplification* of A.

Properties

- Consistent with usual Euclidean vector norm; if $b \in \mathbb{R}^n$, then

$$\|b\| = \sqrt{\lambda_{\mathsf{max}}(b^* b)} = \sqrt{b^* b}$$

- For any x, we have $\|Ax\| \leq \|A\|\|x\|$.
- Scaling: $\|cA\| = |c|\|A\|$
- Triangle inequality: $\|A + B\| \leq \|A\| + \|B\|$.
- Definiteness: $\|A\| = 0 \iff A = 0$.
- Submultiplicative property: $\|AB\| \leq \|A\|\|B\|$.

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.

Note that if $m \neq n$, the matrix Σ is not square; it has the form

$$\Sigma = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ 0 & \cdots & & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & & 0 \end{bmatrix} \quad \text{or} \quad \Sigma = \begin{bmatrix} \sigma_1 & & & 0 & 0 & \dots & 0 \\ & \sigma_2 & & & \vdots & & \vdots \\ & & \ddots & & & & \\ & & & \sigma_m & 0 & \dots & 0 \end{bmatrix}$$

We choose $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$, where $p = \min\{m, n\}$.

Singular value decomposition 2

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, and $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal.

With $U = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix}$ and $V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$, we have

$$A = U \Sigma V^* = \sum_{i=1}^{p} \sigma_i u_i v_i^*$$

- σ_i are the *singular values* of A.
- u_i are the *left singular vectors* of A.
- v_i are the *right singular vectors* of A.

The number of nonzero singular values equals the rank of A.

Singular value decomposition 3

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, and $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal.

We have

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

Hence

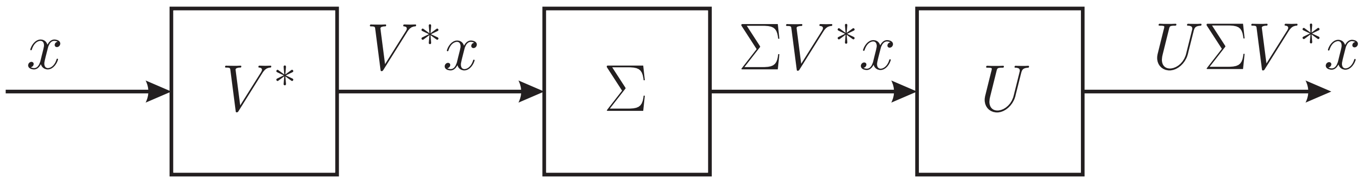
- u_i are the eigenvectors of AA^*
- $\sigma_i = \sqrt{\lambda_i(AA^*)}$ are the eigenvalues of AA^*

Similarly $A^*A = V \Sigma^* U^* U \Sigma V^* = V \Sigma^* \Sigma V^*$ and v_i are the eigenvectors of A^*A, with $\sigma_i = \sqrt{\lambda_i(A^*A)}$ the eigenvalues of A^*A.

If $r = \operatorname{rank}(A)$, then

- $\{u_1, \dots, u_r\}$ are an orthonormal basis for $\operatorname{range}(A)$.
- $\{v_1, \dots, v_r\}$ are an orthonormal basis for $\ker(A)^\perp$.

Linear mapping interpretation of SVD



The SVD decomposes the linear mapping into

- Compute coefficients along directions v_i.
- Scale coefficients by σ_i.
- Generate output along directions u_i.

Note that, unlike the eigen-decomposition, input and output directions are different.

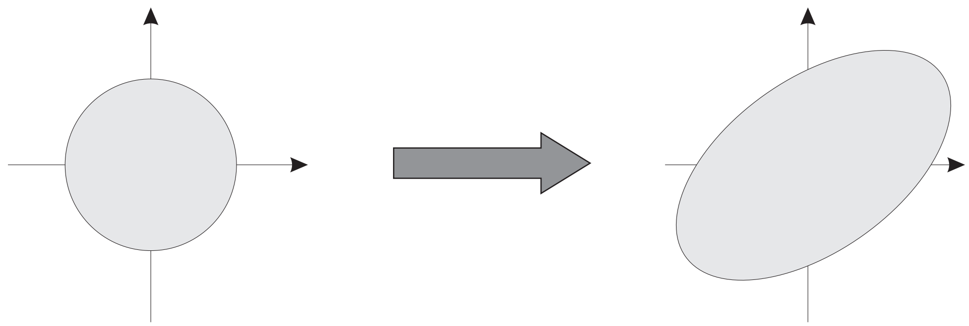
The maximum singular-value, σ_1 gives the norm of A.

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sigma_1$$

The minimum singular-value, σ_p gives the minimum *gain* of the matrix A.

$$\min_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sigma_p$$

Geometric interpretation of SVD



The matrix $A \in \mathbb{R}^{m\times n}$ maps the unit sphere in $\mathbb{R}^n$ to an ellipsoid in $\mathbb{R}^m$.

$$\{x \in \mathbb{R}^n \; ; \; \|x\| = 1\} \quad \to \quad \{y \; ; \; y = Ax, x \in \mathbb{R}^n, \|x\| = 1\}$$

The semi-axes of the ellipse are u_i, with length σ_i.

Note that the ellipse will be degenerate if A is not surjective.

Algebraic interpretation of SVD

The SVD captures the *numerical rank* of a matrix $A \in \mathbb{C}^{m \times n}$.

$$\min\left\{ \|A - B\| \; ; \; B \in \mathbb{C}^{m \times n}, \operatorname{rank}(B) \leq k \right\} = \sigma_{k+1}$$

Theorem: The minimal rank k approximant to A is given by

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^*$$

Hence, if a matrix $A \in \mathbb{R}^{10 \times 10}$ has singular values

$$\sigma_1 = 100, \quad \sigma_2 = 35, \quad \sigma_3 = 10, \quad \sigma_4 = 2$$

and $\sigma_5 \leq 0.00001$, then we might say its *numerical rank* is 4.

Example:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4.01 \end{bmatrix}$$

has singular values $\sigma_1 = 5$, $\sigma_2 = 0.002$. Its optimal rank 1 approximant is

$$A = \begin{bmatrix} 0.9984 & 2.0008 \\ 2.0008 & 4.0096 \end{bmatrix}$$

Proof: We have

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^*$$

hence

$$U^* A_k V = \operatorname{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0).$$

So

$$U^*(A - A_k)V = \operatorname{diag}(0, \ldots, 0, \sigma_k + 1, \ldots, \sigma_p)$$

and hence $\|A - A_k\| = \sigma_{k+1}$.

Now we wish to show that no matrix B can do better. Suppose $\operatorname{rank}(B) = k$ for some B, and let $\{x_1, \ldots, x_{n-k}\}$ be an orthonormal basis for $\ker(B)$. Since $(n-k)+(k+1) > n$,

$$\operatorname{span}\{x_1, \ldots, x_{n-k}\} \cap \operatorname{span}\{v_1, \ldots, v_{k+1}\} \neq \emptyset$$

Let z be a unit vector in this intersection. Then $Bz = 0$, and

$$Az = U\Sigma V^* z = \sum_{i=1}^{k+1} \sigma_i (v_i^* z) u_i \qquad \text{with} \qquad \sum_{i=1}^{k+1} (v_i^* z)^2 = 1$$

hence

$$\|A - B\|^2 \geq \|(A - B)z\|^2 = \|Az\|^2 = \sum_{i=1}^{k+1} \sigma_i^2 (v_i^* z)^2 \geq \sigma_{k+1}^2$$

Example: use of low rank approximants

Suppose $A \in \mathbb{R}^{10000 \times 10000}$ is dense. Then computing the matrix-vector product Ax is computationally expensive; 10^8 multiplications.

But if A has singular values

$$\sigma_1 = 100, \quad \sigma_2 = 35, \quad \sigma_3 = 10, \quad \sigma_4 = 2$$

and $\sigma_k \leq 0.001$ for $k \geq 5$, then the optimal rank 4 approximant is

$$A_4 = \sum_{i=1}^{4} \sigma_i u_i v_i^*$$

Then, let

$$b = A_4 x = 100(v_1^* x)u_1 + 35(v_2^* x)u_2 + 10(v_3^* x)u_3 + 2(v_4^* x)u_4$$

and we have

$$\|Ax - b\| \leq \|A - A_4\| \|x\| \leq 0.001 \|x\|$$

which gives a relative error of 0.1% in 4×10^4 multiplications.

Sensitivity of eigenvalues vs. singular values

Eigenvalues

Suppose

$$A = \begin{bmatrix} 0 & I_9 \\ 0 & 0 \end{bmatrix} \qquad E = \begin{bmatrix} 0 & 0 \\ 10^{-10} & 0 \end{bmatrix}$$

We have

$$\lambda_i(A) = 0 \text{ for all } i \qquad \text{and} \qquad \lambda_{\text{max}}(A + E) = 0.1;$$

A change of order 10^{-10} in A resulted in a change of order 0.1 in its eigenvalues.

The position of the poles of a system can be extremely sensitive to the values of system parameters.

Singular values

Since $\|A\| = \sigma_1(A)$, we know from the triangle inequality that

$$\sigma_1(A + E) \leq \sigma_1(A) + \sigma_1(E)$$

In this case, $\sigma_1(A) = 1$ and $\sigma_1(E) = 10^{-10}$.

Linear matrix inequalities (LMIs)

An inequality of the form

$$F(x) < Q$$

where

- The variable x takes values in a real vector space V.
- The mapping $F : V \to \mathbb{H}^n$ is linear.
- $Q \in \mathbb{H}^n$.

Properties

- A wide variety of control problems can be reduced to a few standard convex optimization problems involving linear matrix inequalities (LMIs).
- The resulting computational problems can be solved *numerically* very efficiently, using *interior-point methods*.
- These algorithms have many important properties, including small computation time, global solutions, provable lower bounds, certificates proving infeasibility, ...
- An LMI formulation often provides an effective solution to a problem.

LMIs in vector form

Every LMI can be represented as

$$F(x) = x_1 F_1 + x_2 F_2 + \cdots + x_m F_m < Q$$

In this case, $x \in \mathbb{R}^m$ and $F_i \in \mathbb{H}^n$.

Example

The inequality

$$\begin{bmatrix} x_1 - 3 & x_1 + x_2 & -1 \\ x_1 + x_2 & x_2 - 4 & 0 \\ -1 & 0 & x_1 \end{bmatrix} < 0$$

is an LMI.

In standard form, we can write this as

$$x_1 \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} + x_2 \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} < \begin{bmatrix} 0 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Semidefinite programming (SDP)

Feasibility problems

Given the LMI

$$F(x) = x_1 F_1 + x_2 F_2 + \cdots + x_m F_m < Q$$

- Find a *feasible point* $x \in \mathbb{R}^m$ such that the LMI is satisfied, or
- determine that there is no such x; that is, that the LMI is *infeasible*.

Linear objective problems

A general problem form is

$$\begin{aligned} \text{minimize} \quad & c^* x \\ \text{subject to} \quad & x_1 F_1 + x_2 F_2 + \cdots + x_m F_m < Q \\ & Ax = b \end{aligned}$$

- Linear cost function
- Equality constraints

LMIs define convex subsets of V

Theorem: The set

$$\mathcal{C} = \Big\{ x \in V \ ; \ F(x) < Q \Big\}$$

is convex.

Proof: We need to show

$$x_1, x_2 \in \mathcal{C}, \quad \theta \in [0,1] \qquad \Longrightarrow \qquad \theta x_1 + (1-\theta) x_2 \in \mathcal{C}$$

Since F is linear,

$$F(\theta x_1 + (1-\theta)x_2) = \theta F(x_1) + (1-\theta) F(x_2) < \theta Q + (1-\theta) Q = Q$$

Alternative proof: The image of a convex set under an affine map is convex.

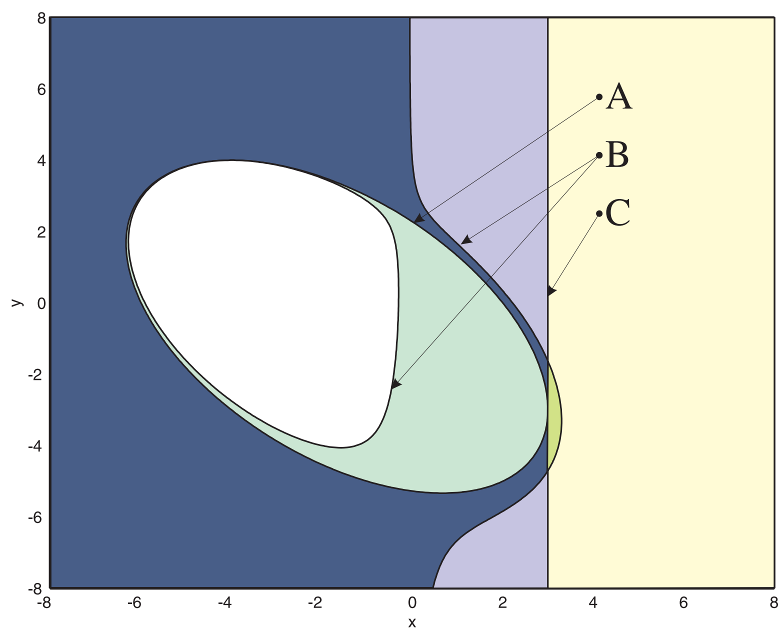
LMIs as polynomial inequalities

Suppose $A \in \mathbb{H}^n$. Let $A_k \in \mathbb{R}^{k\times k}$ be the *submatrix* of A consisting of the first k rows and columns.

Fact: $$A > 0 \quad \Longleftrightarrow \quad \det(A_k) > 0 \text{ for } k = 1, \dots, n.$$

Example:

$$\begin{bmatrix} 3-x_1 & -(x_1+x_2) & 1 \\ -(x_1+x_2) & 4-x_2 & 0 \\ 1 & 0 & -x_1 \end{bmatrix} > 0 \iff \begin{aligned} 3-x_1 > 0 \quad & (C) \\ (3-x_1)(4-x_2)-(x_1+x_2)^2 > 0 \quad & (A) \\ -x_1((3-x_1)(4-x_2)-(x_1+x_2)^2)-(4-x_2) > 0 \quad & (B) \end{aligned}$$



LMIs with matrix variables

Consider the inequality

$$\begin{bmatrix} A^*X + XA & XB \\ B^*X & -I \end{bmatrix} < 0$$

Defining $F : \mathbb{S}^n \to \mathbb{S}^m$ by

$$F(X) = \begin{bmatrix} A^*X + XA & XB \\ B^*X & 0 \end{bmatrix} \qquad \Longrightarrow \qquad F(X) < \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}$$

Notes

- The most common form of LMI in systems and control.
- Easily recognizable.
- Can be more efficient.
- Accepted by software, such as the LMI Control Toolbox.
- Multiple LMIs $G_1(x) < 0, \dots, G_n(x) < 0$ can be converted to one (block-diagonal) LMI

$$\operatorname{diag}(G_1(x), \dots, G_n(x)) < 0$$

LPs can be cast as LMIs

The general linear program

$$\begin{aligned} \text{minimize} \quad & c^* x \\ \text{subject to} \quad & a_1^* x < b_1 \\ & a_2^* x < b_2 \\ & \vdots \\ & a_n^* x < b_n \end{aligned}$$

can be expressed as the SDP

$$\begin{aligned} \text{minimize} \quad & c^* x \\ \text{subject to} \quad & \begin{bmatrix} a_1^* x - b_1 & & & \\ & a_2^* x - b_2 & & \\ & & \ddots & \\ & & & a_n^* x - b_n \end{bmatrix} < 0 \end{aligned}$$

Schur Complements

Recall

$$Q > 0 \text{ and } M - RQ^{-1}R^* > 0 \quad \iff \quad \begin{bmatrix} M & R \\ R^* & Q \end{bmatrix} > 0$$

Example: The matrix $X \in \mathbb{S}^n$ satisfies

$$A^*X - XA + C^*C + XBB^*X < 0$$

if and only if

$$\begin{bmatrix} A^*X + XA + C^*C & XB \\ B^*X & -I \end{bmatrix} < 0$$

This is extremely useful and will reappear often.

Some standard LMIs

Suppose $F_i \in \mathbb{S}^n$, and

$$Z(x) = x_1F_1 + x_2F_2 + \cdots + x_mF_m$$

Matrix norm constraint:

$$\|Z(x)\| < 1 \qquad \iff \qquad \begin{bmatrix} I & Z(x) \\ Z^*(x) & I \end{bmatrix} > 0$$

Matrix norm minimization:

$$\begin{aligned} &\text{minimize} \quad t \\ &\text{subject to} \quad \begin{bmatrix} tI & Z(x) \\ Z^*(x) & tI \end{bmatrix} > 0 \end{aligned}$$

Maximum eigenvalue minimization:

$$\begin{aligned} &\text{minimize} \quad t \\ &\text{subject to} \quad Z(x) - tI < 0 \end{aligned}$$