EE263: Introduction to Linear Dynamical Systems
Review Session 4: Midterm Review

- linear algebra
- left and right inverses
- least squares
- least norm
- Lagrangian

Linear algebra

basic definitions and concepts:

- linearity, vector spaces, subspaces
- inner product, norm

a set of vectors $\{x_1, \ldots, x_k\}$ is

- *independent* if $a_1 x_1 + \cdots + a_k x_k = 0$ holds only when $a_1 = a_2 = \cdots = 0$
- a *basis* for a k-dimensional subspace of $\mathbf{R}^n$ if it is independent
- *orthogonal* if $x_i^T x_j = 0$ for all $i \neq j$
- *normalized* if $\|x_i\| = 1$ for all $i = 1, \ldots, k$

$x_1, \ldots, x_n$ is an *orthonormal basis* for $\mathbf{R}^n$ if it is orthogonal and normalized, then $X = [x_1 \cdots x_n]$ is called orthogonal

the *nullspace* of $A \in \mathbf{R}^{m\times n}$ is defined as $\mathcal{N}(A) = \{x \in \mathbf{R}^n \mid Ax = 0\}$

- $\mathcal{N}(A)$ is the set of vectors that, when multiplied by A, become zero

- if $\mathcal{N}(A) = \{0\}$ (a zero nullspace), then A has a *left inverse* B, which means $BA = I$, and we can unambiguously find x given $y = Ax$ (one-to-one)

the *range* of $A \in \mathbf{R}^{m\times n}$ is defined as $\mathcal{R}(A) = \{Ax \mid x \in \mathbf{R}^n\}$

- the range is all vectors in $\mathbf{R}^m$ that can come out of multiplying A with some vector in $\mathbf{R}^n$. It's the set of all possible outcomes

- if $\mathcal{R}(A) = \mathbf{R}^m$ (all m-vectors are possible outcomes), then A has a *right inverse* B, which means $AB = I$, and we can find at least one solution, $x = By$, to the equation $y = Ax$ (onto) the *rank* of $A \in \mathbf{R}^{m \times n}$ is defined as $\mathbf{Rank}(A) = \dim \mathcal{R}(A)$

- $\mathbf{Rank}(A)$ is less than or equal to $\min(m, n)$
- A is said to be *full-rank* if $\mathbf{Rank}(A) = \min(m, n)$
- $\mathbf{Rank}(A)$ equals the maximum number of independent columns (or rows)

useful properties:

- $\mathbf{Rank}(A) = \mathbf{Rank}(A^T)$
- $\mathbf{Rank}(AB) \leq \min\{\mathbf{Rank}(A), \mathbf{Rank}(B)\}$
- for $A \in \mathbf{R}^{m \times n}$, if $\mathbf{Rank}(A) = r$, then A can be factored as $A = BC$, with $B \in \mathbf{R}^{m \times r}$ and $C \in \mathbf{R}^{r \times n}$
- $\mathcal{R}(A) + \mathcal{N}(A^T) = \mathbf{R}^m$

Left and right inverses

$A \in \mathbf{R}^{m \times n}$ is fat ($m \leq n$) and full rank

- there exists a right inverse B, *i.e.*, $AB = I$
- $A^T(AA^T)^{-1}$ is a right inverse of A

$A \in \mathbf{R}^{m \times n}$ is skinny ($m \geq n$) and full rank

- there exists a left inverse B, *i.e.*, $BA = I$
- $(A^TA)^{-1}A^T$ is a left inverse of A

Least-squares approximate solution

consider a set of linear equations

$$Ax = y$$

where A is skinny and full rank. The least squares approximate solution is $x_{\rm ls} = (A^TA)^{-1}A^Ty$

facts:

- $x_{\rm ls}$ minimizes $\|Ax - y\|^2$
- $A^\dagger = (A^TA)^{-1}A^T$ is called the *pseudo-inverse* of A
- $Ax_{\rm ls}$ is the point in $\mathcal{R}(A)$ that is closest to y, also known as the *projection* of y onto $\mathcal{R}(A)$
- the residual $r = Ax_{\rm ls} - y$ is orthogonal to $\mathcal{R}(A)$

Least norm solution

when $A \in \mathbf{R}^{m \times n}$ is full rank and *fat* $(n \geq m)$, there is definitely a solution (because $\mathcal{R}(A) = \mathbf{R}^m$), and in fact, there are many

facts:

- we can find lots of x's so we have to find a way to choose between them
- one thing to do is to find the *smallest* x that fullfils $y = Ax$
- $x_{\mathrm{ln}} = A^{\dagger} y = A^T (AA^T)^{-1} y$
- $A^{\dagger}$ is a right inverse of A, and it is called the *pseudo-inverse* of A
- it is an *exact* solution of an *undetermined* system $y = Ax$

Example: Point of closest convergence

we have m lines in $\mathbf{R}^n$, described as

$$\mathcal{L}_i = \{p_i + tv_i \mid t \in \mathbf{R}\}, \quad i = 1, \ldots, m,$$

where $p_i \in \mathbf{R}^n$, and $v_i \in \mathbf{R}^n$, with $\|v_i\| = 1$, for $i = 1, \ldots, m$. We define the distance of a point $z \in \mathbf{R}^n$ to a line $\mathcal{L}$ as

$$\mathbf{dist}(z, \mathcal{L}) = \min\{\|z - u\| \mid u \in \mathcal{L}\}$$

we seek a point $z^\star \in \mathbf{R}^n$ that minimizes the sum of the squares of the distances to the lines

$$\sum_{i=1}^{m} \mathbf{dist}(z, \mathcal{L}_i)^2$$

Solution:

- work out an explicit expression for $\mathbf{dist}(z, \mathcal{L}_i)$
- to find this distance we need to solve the simple least-squares problem of minimizing $\|z - p_i - tv_i\|^2$ over $t \in \mathbf{R}$
- the optimal t is given by $t^\star = v_i^T(z - p_i)$, so we have

$$\mathbf{dist}(z, \mathcal{L}_i) = \|z - p_i - t^\star v_i\| = \|(I - v_i v_i^T)(z - p_i)\|$$

- notice: $I - v_i v_i^T$ is projection onto the orthogonal complement of the line through the origin in the direction v_i, *i.e.*, projection onto the plane with normal vector v_i
- standard least-squares problem

$$A = \begin{bmatrix} I - v_1 v_1^T \\ \vdots \\ I - v_m v_m^T \end{bmatrix}, \quad b = \begin{bmatrix} (I - v_1 v_1^T)p_1 \\ \vdots \\ (I - v_m v_m^T)p_m \end{bmatrix},$$

$$\sum_{i=1}^{m} \mathbf{dist}(z, \mathcal{L}_i)^2 = \|Az - b\|^2$$

- assuming A is full rank, the solution is

$$z^\star = (A^T A)^{-1} A^T b = \left(mI - \sum_{i=1}^{m} v_i v_i^T \right)^{-1} \sum_{i=1}^{m} (p_i - v_i v_i^T p_i)$$

Example: Middle inverse

consider the matrix equation

$$AXB = I$$

where $A \in \mathbf{R}^{n\times p}$, $B \in \mathbf{R}^{q\times n}$, and $X \in \mathbf{R}^{p\times q}$. Goal: determine X that satisfies this equation

Solution:

when can we do it?

- when A and B are full rank, in particular, when A is fat and B is skinny
- $\mathcal{N}(B)$ cannot contain a non-zero vector, likewise, $\mathcal{N}(A^T)$ cannot contain a non-zero vector

how?

- we can find matrices C and D such that $CA^T = I$ and $DB = I$
- we then have that

$$AC^T DB = (CA^T)^T (DB) = I$$

- $X = C^T D$

if X exists, then in general,

$$X = A^T(AA^T)^{-1}(B^TB)^{-1}B^T$$

- $A^T(AA^T)^{-1}$ is a right inverse of A
- $(B^TB)^{-1}B^T$ is a left inverse of B

Lagrangian

Lagrangian method provides a procedure to solve optimization problems with equality constraints

standard form problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

variable $x \in \mathbf{R}^n$ and arbitrary functions $f : \mathbf{R}^n \to \mathbf{R}$ and $h_i : \mathbf{R}^n \to \mathbf{R}$

Lagrangian: $L : \mathbf{R}^n \times \mathbf{R}^p \to \mathbf{R}$ is defined as

$$L(x, \lambda) = f(x) + \sum_{i=1}^{p} \lambda_i h_i(x)$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $h_i(x) = 0$

- optimality conditions are

$$\nabla_x L = 0, \qquad \nabla_\lambda L = 0$$

$\nabla_x L$ and $\nabla_\lambda L$ are in general nonlinear functions of x and λ

- in cases when f is quadratic and h_i are affine, $\nabla_x L$ and $\nabla_\lambda L$ are affine
- in the least norm problem, we have

$$\nabla_x L = 2x + A^T\lambda = 0, \qquad \nabla_\lambda L = Ax - y = 0$$

we can solve x and λ by a linear equation

$$\begin{bmatrix} 2I & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}$$