Linear algebra: intermediate-advanced topics[1]

NOTATIONS.

(1) $\mathfrak{i}$ — the imaginary unit: $\mathfrak{i}^2 = -1$.
(2) $A^\top$ — the transpose of the matrix A.
(3) A^* — the transpose of the matrix A and complex conjugation of the entries. `Hermitian matrix` $A = A^*$.
(4) I_n — the identity matrix of size n.
(5) S_n — the symmetric group: the collection of all permutations (bijections) of the set $(1, 2, \ldots, n)$;

The determinant of a square n-by-n matrix $A = (a_{ij})$, with real or complex entries, is the number

$$\det(A) = |A| = \sum_{\sigma \in S_n} (-1)^{|\sigma|} a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)},$$

where, for $\sigma \in S_n$, $|\sigma|$ is the number of times a bigger number comes in front of the smaller number in the sequence $(\sigma(1), \sigma(2), \ldots, \sigma(n))$.

The permanent of a square n-by-n matrix $A = (a_{ij})$, with real or complex entries, is the number

$$\text{perm}(A) = |A| = \sum_{\sigma \in S_n} a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}.$$

Some applications for real matrices: sum of weights of all cycle-covers of the weighted digraph with adjacency matrix A; sum of the weights of all perfect matchings of the weighted bipartite graph with adjacency matrix A.

The Pfaffian of a $2n$-by-$2n$ *skew-symmetric* matrix A $[A^\top = -A]$, with real or complex entries, is the number

$$\text{pf}(A) = \frac{1}{2^n n!} \sum_{\sigma \in S_{2n}} (-1)^{|\sigma|} a_{\sigma(1),\sigma(2)} a_{\sigma(3),\sigma(4)} \cdots a_{\sigma(2n-1),\sigma(2n)};$$

for odd-dimensional skew-symmetric matrices, we set $\text{pf}(A) = 0$.

Theorem. If A is a $2n$-by-$2n$ skew-symmetric matrix, then $\text{pf}(A^\top) = (-1)^n \text{pf}(A)$ and $\det(A) = \big(\text{pf}(A)\big)^2$. If B is a square matrix of the same size as A, then $\text{pf}(B^\top AB) = \det(B)\text{pf}(A)$. In particular, with $B = A^m$, $\text{pf}(A^{2m+1}) = (-1)^{mn}\big(\text{pf}(A)\big)^{2m+1}$.

Properties of the trace

$$\text{Tr}(AB) = \text{Tr}(BA),\ \text{Tr}(ABC) = \text{Tr}(CAB) = \text{Tr}(BCA),\ \text{Tr}\big(p(A)\big) = \sum_k p(\lambda_k),\ \det\big(e^A\big) = e^{\text{Tr}(A)},$$

where $p = p(x)$ is a polynomial. In general, $\text{Tr}(ABC) \neq \text{Tr}(BAC)$ For *Hermitian* matrices A, B, $|\text{Tr}(AB)| \leq \|A\|_p^{(s)} \|B\|_q^{(s)}$, where $\|\cdot\|_p^{(s)}$ is the *Schatten* p-norm and $(1/p) + (1/q) = 1$.

Cayley-Hamilton Theorem. If $p_A(x) = \det(A - xI_n)$ is the characteristic polynomial of the square matrix A, then $p_A(A) = 0$, the zero matrix. *The main consequence* is that, when it comes to powers of an n-by-n matrix, only the first n power matter, *at most*[2], and all powers higher than n can be expressed in terms of the lower powers.

Vectorization operation $\text{vec} : \mathbb{K}^{m\times n} \mapsto \mathbb{K}^{mn\times 1}$, $\mathbb{K}$ is either $\mathbb{R}$ or $\mathbb{C}$. A matrix is turned into a vector by stacking the *columns* of the matrix, going from left to right, into one column vector.

The Kronecker product $A \otimes B$ of m-by-n matrix A and p-by-q matrix B is a matrix of size mp-by-nq. The matrix has a *block* structure, with $a_{ij}B$ being the block in position i, j. Similar to the usual matrix product, this operation is bi-linear, associative, and non-commutative. By direct computation, the matrix equation $AXB^\top = C$ can be written in the matrix-vector form for the unknown vector $\vec{X} = \text{vec}(X)$ as $(B \otimes A)\vec{X} = \text{vec}(C)$.

---

[1]Sergey Lototsky, USC, version of May 18, 2023

[2]the `minimal` polynomial of a matrix might have degree lower than n

Matrix norm is a norm $\|\cdot\|$ on the linear space of square matrices, with additional requirement $\|AB\| \leq \|A\|\,\|B\|$. Main examples:

- `Schatten` p-norm, $1 \leq p \leq +\infty$ is the L_p norm of the sequence of the *singular values*.
- `Frobenius` or Hilbert-Schmidt norm $\left(\mathrm{Tr}(A^*A)\right)^{1/2}$ is the same as the Euclidean norm of $\mathrm{vec}(A)$ and is the same as the Schatten 2-norm.
- `Spectral` norm or the *default* `operator norm` is largest singular value of A and also the operator norm of A as the map between the corresponding *Euclidean* spaces.
- `Trace` (class) norm or *nuclear* norm is the sum of singular values, that is, Schatten 1-norm.

Matrix decompositions/factorizations:

- `Cholesky` $A = LL^*$ for a Hermitian matrix A.
- `Jordan` $A = BJB^{-1}$ for any square matrix; J is the *Jordan canonical form* of A.
- `LU` $A = LU$ for *some* square matrices.
- `Polar` $A = Q(A^*A)^{1/2}$ for any square matrix A.
- `QR` $A = QR$ for any square matrix A, with Q unitary and R upper triangular.
- `Schur` $A = QUQ^*$ for any square matrix A.
- `Spectral` $A = QDQ^*$ for *normal* matrix ($AA^* = A^*A$).
- `SVD` (singular value) $A = UDV^*$ for any matrix A.

Some advanced notions, informally summarized. A `Lie algebra` is an algebra [linear space with a bi-linear form] in which the bi-linear form, known as the `Lie bracket`, is anti-commutative and satisfies the *Jacobi identity* [think cross product in $\mathbb{R}^3$ or commutator for matrices]. A (topological n-dimensional) `manifold` is a Hausdorff space [different points have disjoint neighborhoods — the most common of at least 10 possible "separation conditions"] where the neighborhood of every point is homeomorphic [bijection, continuous in both directions] to an open subset of $\mathbb{R}^n$; for a `smooth` manifold, those local homeomorphisms are, in a sense, infinitely differentiable [extra constructions are necessary to ensure that we deal with functions from $\mathbb{R}^n$ to itself]. A `Lie group` is a set that is both a group and a smooth manifold, and the group operations are consistent with the smooth structure of the manifold [the shortest formal definition uses language of *category theory*]. A `Lie algebra of a Lie group` is the tangent space at the identity element with the corresponding *Lie bracket* as the bi-linear form. For *matrix Lie groups*, the corresponding Lie bracket is the commutator $[A, B] = AB - BA$.

Special collections of matrices

- $GL_n(\mathbb{K})$ — `general linear group` (change of basis): invertible n-by-n matrices with real ($\mathbb{K} = \mathbb{R}$) or complex ($\mathbb{K} = \mathbb{C}$) entries.
- $SL_n(\mathbb{K})$ — `special linear group` (change of basis that preserves orientation and volume): matrices from $GL_n(\mathbb{K})$ with determinant equal to 1.
- $O(n)$ — `orthogonal group` (isometries of $\mathbb{R}^n$): $\{A \in GL_n(\mathbb{R}) : A^\top A = I_n\}$.
- $SO(n)$ — `special orthogonal group` (rotations of $\mathbb{R}^n$): $O(n) \bigcap SL_n(\mathbb{R})$.
- $U(n)$ — `unitary group` (isometries of $\mathbb{C}^n$): $\{A \in GL_n(\mathbb{C}) : A^*A = I_n\}$.
- $SU(n)$ — `special unitary group`: $U(n) \bigcap SL_n(\mathbb{C})$
- $\mathfrak{sl}_n(\mathbb{K})$ — `the Lie algebra` of $SL_n(\mathbb{K})$: matrices with zero trace.
- $\mathfrak{o}(n)$ — `the Lie algebra` of $O(n)$: skew-symmetric matrices $A = -A^\top$.
- $\mathfrak{so}(n)$ — `the Lie algebra` $SO(n)$: $\mathfrak{sl}_n(\mathbb{R}) \bigcap \mathfrak{o}(n)$.
- $\mathfrak{u}(n) = \{A : A^* = -A\}$, $\mathfrak{su}(n) = \mathfrak{u}(n) \bigcap \mathfrak{sl}_n(\mathbb{C})$.

The *deep underlying relation*: if $A \in \mathfrak{gl}_n(\mathbb{K})$ (that is, an *arbitrary* matrix), then $e^A \in GL_n(\mathbb{K})$.