2023-01-30

1 Matrices and mappings

A matrix represents a mapping between two vector spaces. That is, if $L : \mathcal{V} \rightarrow \mathcal{W}$ is a linear map, then the associated matrix A with respect to bases V and W satisfies $A = W^{-1}LV$. The same linear mapping corresponds to different matrices depending on the choices of basis. But matrices can represent several other types of mappings as well. Over the course of this class, we will see several interpretations of matrices:

- **Linear maps**. A map $L : \mathcal{V} \rightarrow \mathcal{W}$ is linear if $L(x + y) = Lx + Ly$ and $L(\alpha x) = \alpha Lx$. The corresponding matrix is $A = W^{-1}LV$.

- **Linear operators**. A linear map from a space to itself ($L : \mathcal{V} \rightarrow \mathcal{V}$) is a linear operator. The corresponding (square) matrix is $A = V^{-1}LV$.

- **Bilinear forms**. A map $a : \mathcal{V} \times \mathcal{W} \rightarrow \mathbb{R}$ (or $\mathbb{C}$ for complex spaces) is bilinear if it is linear in both slots: $a(\alpha u + v, w) = \alpha a(u, w) + a(v, w)$ and $a(v, \alpha u + w) = \alpha a(v, u) + a(v, w)$. The corresponding matrix has elements $A_{ij} = a(v_i, w_j)$; if $v = Vc$ and $w = Wd$ then $a(v, w) = d^T Ac$.

  We call a bilinear form on $\mathcal{V} \times \mathcal{V}$ *symmetric* if $a(v, w) = a(w, v)$; in this case, the corresponding matrix A is also symmetric ($A = A^T$). A symmetric form and the corresponding matrix are called *positive semi-definite* if $a(v, v) \geq 0$ for all v. The form and matrix are *positive definite* if $a(v, v) > 0$ for any $v \neq 0$.

  A *skew-symmetric* matrix ($A = -A^T$) corresponds to a skew-symmetric or anti-symmetric bilinear form, i.e. $a(v, w) = -a(w, v)$.

- **Sesquilinear forms**. A map $a : \mathcal{V} \times \mathcal{W} \rightarrow \mathbb{C}$ (where $\mathcal{V}$ and $\mathcal{W}$ are complex vector spaces) is sesquilinear if it is linear in the first slot and the conjugate is linear in the second slot: $a(\alpha u + v, w) = \alpha a(u, w) + a(v, w)$ and $a(v, \alpha u + w) = \bar{\alpha} a(v, u) + a(v, w)$. The matrix has elements $A_{ij} = a(v_i, w_j)$; if $v = Vc$ and $w = Wd$ then $a(v, w) = d^* Ac$.

  We call a sesquilinear form on $\mathcal{V} \times \mathcal{V}$ *Hermitian* if $a(v, w) = a(w, v)$; in this case, the corresponding matrix A is also Hermitian ($A = A^*$). A

Hermitian form and the corresponding matrix are called *positive semi-definite* if $a(v, v) \geq 0$ for all v. The form and matrix are *positive definite* if $a(v, v) > 0$ for any $v \neq 0$.

A *skew-Hermitian* matrix ($A = -A^*$) corresponds to a skew-Hermitian or anti-Hermitian bilinear form, i.e. $a(v, w) = -a(w, v)$.

- **Quadratic forms**. A quadratic form $\phi : \mathcal{V} \rightarrow \mathbb{R}$ (or $\mathbb{C}$) is a homogeneous quadratic function on $\mathcal{V}$, i.e. $\phi(\alpha v) = |\alpha|^2 \phi(v)$ for which the map $b(v, w) = \phi(v + w) - \phi(v) - \phi(w)$ is bilinear. Any quadratic form on a finite-dimensional space can be represented as $c^* A c$ where c is the coefficient vector for some Hermitian matrix A. The formula for the elements of A given ϕ is left as an exercise.

We care about linear maps and linear operators almost everywhere, and most students come out of a first linear algebra class with some notion that these are important. But apart from very standard examples (inner products and norms), many students have only a vague notion of what a bilinear form, sesquilinear form, or quadratic form might be. Bilinear forms and sesquilinear forms show up when we discuss large-scale solvers based on projection methods. Quadratic forms are important in optimization, physics (where they often represent energy), and statistics (e.g. for understanding variance and covariance).

1.1 Matrix norms

The space of matrices forms a vector space; and, as with other vector spaces, it makes sense to talk about norms. In particular, we frequently want norms that are *consistent* with vector norms on the range and domain spaces; that is, for any w and v, we want

$$w = Av \implies \|w\| \leq \|A\| \|v\|.$$

One "obvious" consistent norm is the *Frobenius norm*,

$$\|A\|_F^2 = \sum_{i,j} a_{ij}^2.$$

Even more useful are *induced norms* (or *operator norms*)

$$\|A\| = \sup_{v \neq 0} \frac{\|Av\|}{\|v\|} = \sup_{\|v\|=1} \|Av\|.$$

The induced norms corresponding to the vector 1-norm and ∞-norm are

$$\|A\|_1 = \max_j \sum_i |a_{ij}| \quad \text{(max abs column sum)}$$
$$\|A\|_\infty = \max_i \sum_j |a_{ij}| \quad \text{(max abs row sum)}$$

The norm induced by the vector Euclidean norm (variously called the matrix 2-norm or the spectral norm) is more complicated.

The Frobenius norm and the matrix 2-norm are both *orthogonally invariant* (or *unitarily invariant* in a complex vector space. That is, if Q is a square matrix with $Q^* = Q^{-1}$ (an orthogonal or unitary matrix) of the appropriate dimensions

$$\begin{aligned} \|QA\|_F &= \|A\|_F, & \|AQ\|_F &= \|A\|_F, \\ \|QA\|_2 &= \|A\|_2, & \|AQ\|_2 &= \|A\|_2. \end{aligned}$$

This property will turn out to be frequently useful throughout the course.

1.2 Decompositions and canonical forms

Matrix decompositions (also known as *matrix factorizations*) are central to numerical linear algebra. We will get to know six such factorizations well:

- $PA = LU$ (a.k.a. Gaussian elimination). Here L is unit lower triangular (triangular with 1 along the main diagonal), U is upper triangular, and P is a permutation matrix.

- $A = LL^*$ (a.k.a. Cholesky factorization). Here A is Hermitian and positive definite, and L is a lower triangular matrix.

- $A = QR$ (a.k.a. QR decomposition). Here Q has orthonormal columns and R is upper triangular. If we think of the columns of A as a basis, QR decomposition corresponds to the Gram-Schmidt orthogonalization process you have likely seen in the past (though we rarely compute with Gram-Schmidt).

- $A = U\Sigma V^*$ (a.k.a. the singular value decomposition or SVD). Here U and V have orthonormal columns and Σ is diagonal with non-negative entries.

- $A = Q\Lambda Q^*$ (a.k.a. symmetric eigendecomposition). Here A is Hermitian (symmetric in the real case), Q is orthogonal or unitary, and Λ is a diagonal matrix with real numbers on the diagonal.
- $A = QTQ^*$ (a.k.a. Schur form). Here A is a square matrix, Q is orthogonal or unitary, and T is upper triangular (or nearly so).

The last three of these decompositions correspond to *canonical forms* for abstract operators. That is, we can view these decompositions as finding bases in which the matrix representation of some operator or form is particularly simple. More particularly:

- **SVD**: For any linear mapping $L : \mathcal{V} \to \mathcal{W}$, there are orthonormal bases for the two spaces such that the corresponding matrix is diagonal
- **Symmetric eigendecomposition**: For any Hermitian sesquilinear map on an inner product space, there is an orthonormal basis for the space such that the matrix representation is diagonal.
- **Schur form**: For any linear operator $L : \mathcal{V} \to \mathcal{V}$, there is an orthonormal basis for the space such that the matrix representation is upper triangular. Equivalently, if $\{u_1, \ldots, u_n\}$ is the basis in question, then $\operatorname{sp}(\{u_j\}_{j=1}^k)$ is an *invariant subspace* for each $1 \leq k \leq n$.

The Schur form turns out to be better for numerical work than the Jordan canonical form that you should have seen in an earlier class. We will discuss this in more detail when we discuss eigenvalue problems.

1.3 The SVD and the 2-norm

The singular value decomposition is useful for a variety of reasons; we close off the lecture by showing one such use.

Suppose $A = U\Sigma V^*$ is the singular value decomposition of some matrix. Using orthogonal invariance (unitary invariance) of the 2-norm, we have

$$\|A\|_2 = \|U^* A V\|_2 = \|\Sigma_2\|,$$

i.e.

$$\|A\|_2 = \max_{\|v\|^2=1} \frac{\sum_j \sigma_j |v_j|^2}{\sum |v_j|^2}.$$

That is, the spectral norm is the largest weighted average of the singular values, which is the same as just the largest singular value.

The small singular values also have a meaning. If A is a square, invertible matrix then

$$\|A^{-1}\|_2 = \|V\Sigma^{-1}U^*\|_2 = \|\Sigma_{-1}\|_2,$$

i.e. $\|A^{-1}|_2$ is the inverse of the smallest singular value of A.

The smallest singular value of a nonsingular matrix A can also be interpreted as the "distance to singularity": if σ_n is the smallest singular value of A, then there is a matrix E such that $\|E\|_2 = \sigma_n$ and $A + E$ is singular; and there is no such matrix with smaller norm.

These facts about the singular value decomposition are worth pondering, as they will be particularly useful in the next lecture when we ponder sensitivity and conditioning.