EECS127 Course Notes

Anmol Parande

Spring 2021 - Professor Laurent El Ghaoui

Disclaimer: These notes reflect 127 when I took the course (Spring 2021). They may not accurately reflect current course content, so use at your own risk. If you find any typos, errors, etc, please raise an issue on the GitHub repository.

Contents

1 Math Overview

1.1 Vectors

Definition 1 *An affine set is one of the form $\mathcal{A} = \{\boldsymbol{x} \in \mathcal{X} : \ \boldsymbol{x} = \boldsymbol{v} + \boldsymbol{x_o}, \ \boldsymbol{v} \in \mathcal{V}\}$ where $\mathcal{V}$ is a subspace of a vector space $\mathcal{X}$ and x_0 is a given point.*

Notice that by definition 1, a subspace is simply an affine set containing the origin. Also notice that the dimension of an affine set $\mathcal{A}$ is the same as the dimension of $\mathcal{V}$. For a given vector space, we can define a function which maps that vector to a real number.

Definition 2 *A norm on the vector space $\mathcal{X}$ is a function $\|\cdot\| : \mathcal{X} \to \mathbb{R}$ which satisfies $\|\boldsymbol{x}\| \geq 0$ with equality if and only if $\boldsymbol{x} = \boldsymbol{0}$, $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$, and $\|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|$ for any scalar α.*

Definition 3 *The l_p norms are defined by*

$$\|\boldsymbol{x}\|_p = \left(\sum_{k=1}^{n} |x_k|^p \right)^{\frac{1}{p}}, \ 1 \leq p \leq \infty$$

Notice that for $p = 2$, we recover the Euclidean norm, and in the limit as $p \to \infty$, $\|\boldsymbol{x}\|_\infty = \max_k |x_k|$.

Definition 4 *An inner product on real vector space is a function that maps $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ to a non-negative scalar, is distributive, is commutative, and $\langle \boldsymbol{x}, \boldsymbol{x}, \rangle = 0 \Leftrightarrow \boldsymbol{x} = 0$.*

Inner products induce a norm $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$. In $\mathbb{R}^n$, the standard inner product is $\boldsymbol{x}^T \boldsymbol{y}$. The angle bewteen two vectors is given by

$$\cos\theta = \frac{\boldsymbol{x}^T \boldsymbol{y}}{\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2}$$

There are two other properties which use the standard inner product.

Theorem 1 (Cauchy-Schwarz Inequality)

$$|\boldsymbol{x}^T \boldsymbol{y}| \leq \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2$$

Theorem 2 (Holder Inequality)

$$|\boldsymbol{x}^T \boldsymbol{y}| \leq \sum_{k=1}^{n} |x_k y_k| \leq \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q, \ p, q \geq 1 \ s.t \ p^{-1} + q^{-1} = 1.$$

Notice that theorem 2 generalizes theorem 1.

1.1.1 Projection

The idea behind projection is to find the closest point in a set closest (with respect to particular norm) to a given point.

Definition 5 *Given a vector $\boldsymbol{x}$ in inner product space $\mathcal{X}$ and a subset $S \subseteq \mathcal{X}$, the projection of $\boldsymbol{x}$ onto S is given by*

$$\Pi_S(\boldsymbol{x}) = \operatorname*{argmin}_{\boldsymbol{y} \in S} \|\boldsymbol{y} - \boldsymbol{x}\|$$

where the norm is the one induced by the inner product.

Theorem 3 *There exists a unique vector $\boldsymbol{x}^* \in S$ which solves*

$$\min_{\boldsymbol{y} \in S} \|\boldsymbol{y} - \boldsymbol{x}\|.$$

It is necessary and sufficient for $\boldsymbol{x}^*$ to be optimal that $(\boldsymbol{x} - \boldsymbol{x}^*) \perp S$. The same condition applies when projecting onto an affine set.

1.2 Functions

We consider functions to be of the form $f : \mathbb{R}^n \to \mathbb{R}$. By contrast, a map is of the form $f : \mathbb{R}^n \to \mathbb{R}^m$. The components of the map f are the scalar valued functions f_i that produce each component of a map.

Definition 6 *The graph of a function f is the set of input-output pairs that f can attain.*

$$\left\{(x, f(x)) \in \mathbb{R}^{n+1} : \ x \in \mathbb{R}^n\right\}$$

Definition 7 *The epigraph of a function is the set of input-output pairs that f can achieve and anything above.*

$$\left\{(x, t) \in \mathbb{R}^{n+1} : \ \boldsymbol{x} \in \mathbb{R}^{n+1}, \ t \geq f(x)\right\}$$

Definition 8 *The t-level set is the set of points that achieve exactly some value of f.*

$$\{\boldsymbol{x} \in \mathbb{R}^n : \ f(x) = t\}$$

Definition 9 *The t-sublevel set of f is the set of points achieving at most a value t.*

$$\{x \in \mathbb{R}^n : \ f(x) \le t\}$$

Theorem 4 *A function is linear if and only if it can be expressed as $f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + b$ for some unique pair $(\boldsymbol{a}, b)$.*

An affine function is linear when $b = 0$. A hyperplane is simply a level set of a linear function.

Definition 10 *The half-spaces are the regions of space which a hyper-plane separates.*

$$H_{-} = \{x : \boldsymbol{a}^T\boldsymbol{x} \le b\} \qquad H_{+} = \{x : \boldsymbol{a}^T\boldsymbol{x} > b\}$$

Definition 11 *The gradient of a function at a point x where f is differentiable is a column vector of first derviatives of f with respsect to the components of $\boldsymbol{x}$*

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

The gradient is perpendicular to the level sets of f and points from a point $\boldsymbol{x}_0$ to higher values of the function. In other words, it is the direction of steepest increase.

1.3 Matrices

Matrices define a linear map between an input space and an output space. Any linear map $f : \mathbb{R}^n \to \mathbb{R}^m$ can be represented by a matrix.

Theorem 5 (Fundamental Theorem of Linear Algebra) *For any matrix $A \in \mathbb{R}^{m \times n}$,*

$$\mathcal{N}(A) \oplus \mathcal{R}(A^T) = \mathbb{R}^n \qquad \mathcal{R}(A) \oplus \mathcal{N}(A^T) = \mathbb{R}^m.$$

Similar to vectors, matrices can also have norms.

Definition 12 *A function $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ is a matrix norm if*

$$f(A) \geq 0 \quad f(A) = 0 \Leftrightarrow A = 0 \quad f(\alpha A) = |\alpha| f(A) \quad f(A+B) \leq f(A) + f(B)$$

Definition 13 *The Froebenius norm is the l_2 norm applied to all elements of the matrix.*

$$\|A\|_F = \sqrt{traceAA^T} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}$$

Definition 14 *The operator norms is defined as*

$$\|A\|_p = \max_{\boldsymbol{u}\neq 0} \frac{\|A\boldsymbol{u}\|_p}{\|u\|_p}$$

The operator norms characterize the maximum input-output gain of the matrix A measured relative to some l_p norm.

1.3.1 QR Factorization

Definition 15 *The QR factorization matrix are the orthogonal matrix Q and the upper triangular matrix R such that $A = QR$*

An easy way to find the QR factorization of a matrix is to apply Graham Schmidt to the columns of the matrix and express the result in matrix form. Suppose that our matrix A is full rank (i.e its columns $\boldsymbol{a}_i$ are linearly independent) and we have applied Graham-Schmidt to columns $\boldsymbol{a}_{i+1}\cdots\boldsymbol{a}_n$ to get orthogonal vectors $\boldsymbol{q}_{i+1}\cdots\boldsymbol{q}_n$. Continuing the procedure, the ith orthogonal vector $\boldsymbol{q}_i$ is

$$\tilde{\boldsymbol{q}}_i = \boldsymbol{a}_i - \sum_{k=i+1}^{n}(\boldsymbol{q}_k^T\boldsymbol{a}_k)\boldsymbol{q}_k \qquad \boldsymbol{q}_i = \frac{\tilde{\boldsymbol{q}}_i}{\|\tilde{\boldsymbol{q}}_i\|_2}.$$

If we re-arrange this, to solve for $\boldsymbol{a}_i$, we see that

$$\boldsymbol{a}_i = \|\tilde{\boldsymbol{q}}_i\|_2\boldsymbol{q}_i + \sum_{k=i+1}^{n}(\boldsymbol{q}_k^T\boldsymbol{a}_k)\boldsymbol{q}_k.$$

Putting this in matrix form, we can see that

$$\begin{bmatrix} | & | & & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{nn} \end{bmatrix} \qquad r_{ij} = \boldsymbol{a}_i^T\boldsymbol{q}_{\boldsymbol{j}}, r_{ii} = \|\tilde{\boldsymbol{q}}_i\|_2.$$

1.3.2 Symmetric Matrices

In Linear Algebra, symmetrics matrices are nice because they have numerous useful properties.

Theorem 6 (Spectral Theorem) *Any symmetric matrix is orthogonally similar to a real diagonal matrix.*

$$A = A^T \implies A = U\Lambda U^T = \sum_i \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T, \quad \|\boldsymbol{u}\| = 1, \quad \boldsymbol{u}_i^T \boldsymbol{u}_j = 0 \ (i \neq j)$$

Let $\lambda_{min}(A)$ be the smallest eigenvalue of symmetric matrix A and $\lambda_{max}(A)$ be the largest eigenvalue.

Definition 16 *The Rayleigh Quotient is*

$$\frac{\boldsymbol{x}^T A \boldsymbol{x}}{\|\boldsymbol{x}\|^2}$$

for $\boldsymbol{x} \neq \mathbf{0}$ and it is always true that

$$\lambda_{min}(A) = \frac{\boldsymbol{x}^T A \boldsymbol{x}}{\|\boldsymbol{x}\|^2} \leq \lambda_{max}(A)$$

Suppose we wanted to find how much a linear operation A "magnifies" its input. We can characterize that by the largest eigenvalue of $A^T A$

Definition 17 *The spectral norm is the maximum gain of a matrix with respect to the spectral norm.*

$$\|A\|_2 = \max_{\boldsymbol{x} \neq \mathbf{0}} = \frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \sqrt{\lambda_{max}(A^T A)}$$

Two special types of symmetric matrices are those with non-negative eigenvalues.

Definition 18 *A symmetric matrix is positive semi-definite if $\boldsymbol{x}^T A \boldsymbol{x} \geq 0 \implies \lambda_{min}(A) \geq 0$.*

Definition 19 *A symmetric matrix is poitive definite if $\boldsymbol{x}^T A \boldsymbol{x} > 0 \implies \lambda_{min}(A) > 0$.*

These matrices are important because they often have very clear geometric structures. For example, an ellipsoid in multi-dimensional space can be defined as the set of points

$$\mathcal{E} = \{x \in \mathbb{R}^m : \ \boldsymbol{x}^T P^{-1} \boldsymbol{x} \leq 1\}$$

where P is a positive definite matrix. The eigenvectors of P give the principle axes of this ellipse, and $\sqrt{\lambda}$ are the semi-axis lengths.