16-485: Visual Navigation for Autonomous Vehicles (VNAV) **Fall 2020**

Recitation 1: Mathematical Preliminaries

Lecturer: Heng Yang *Scribes: Heng Yang*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor(s).*

This recitation serves as a recap of useful mathematical definitions and tools for VNAV 2020, and for research in robotics and computer vision in general. It covers basic linear algebra (Section 1.1) and matrix calculus (Section 1.2). These notes have benefited from [3, 2].

Notations. We use lowercase characters (*e.g.,* $s \in \mathbb{R}, \mathbb{C}$) to denote real and complex scalars, bold lowercase characters (*e.g.,* $\boldsymbol{v} \in \mathbb{R}^n, \mathbb{C}^n$) for real and complex vectors, and bold uppercase characters (*e.g.,* $\boldsymbol{M} \in \mathbb{R}^{m \times n}, \mathbb{C}^{m \times n}$) for real and complex matrices. v_i denotes the i-th scalar entry of vector $\boldsymbol{v}$, and M_{ij} denotes the i-th row and j-th column scalar entry of matrix $\boldsymbol{M}$. For a square matrix $\boldsymbol{M} \in \mathbb{R}^{n \times n}$, $\operatorname{tr}(\boldsymbol{M}) \doteq \sum_{i=1}^{n} M_{ii}$ denotes the *trace* of $\boldsymbol{M}$, and $\det(\boldsymbol{M})$ denotes the *determinant* of $\boldsymbol{M}$. For any matrix $\boldsymbol{M} \in \mathbb{R}^{m \times n}$, denote $\operatorname{vec}(\boldsymbol{M}) \in \mathbb{R}^{mn}$ as the column-wise *vectorization* of $\boldsymbol{M}$ by vertically stacking its columns. We use $\mathcal{S}^n$ to denote the set of real *symmetric* matrices of size $n \times n$. For any vector $\boldsymbol{v} \in \mathbb{R}^n$, $\operatorname{diag}(\boldsymbol{v})$ creates a diagonal matrix $\boldsymbol{V} \in \mathcal{S}^n$ with diagonal entries $V_{ii} = v_i, i = 1, \ldots, n$.

1.1 Linear Algebra

1.1.1 Norms

Inner Product. The standard inner product on $\mathbb{R}^n$, the set of n-dimensional real vectors, is defined as:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^{\mathsf{T}} \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i. \tag{1.1}$$

The standard inner product on $\mathbb{R}^{m \times n}$, the set of $m \times n$ real matrices, is defined as:

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \operatorname{tr}\left(\boldsymbol{X}^{\mathsf{T}} \boldsymbol{Y}\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij} Y_{ij}. \tag{1.2}$$

Vector Norms. Let us first introduce the definition of a general vector norm in $\mathbb{R}^n$.

Definition 1 (General Norm)**.** *A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is called a* norm *if f satisfies:*

(i) Nonnegativity: $f(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$;

(ii) Definiteness: $f(\boldsymbol{x}) = 0$ if and only if $\boldsymbol{x} = \boldsymbol{0}$;

(iii) Nonnegative homogeneity: $f(t\boldsymbol{x}) = |t| f(\boldsymbol{x})$ for all $t \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^n$;

(iv) Triangle inequality: $f(\boldsymbol{x} + \boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y})$, for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.

When f satisfies Definition 1, f is called a norm function and typically denoted as $\|\cdot\|$. We use $\|\boldsymbol{x}\|_p$ to denote the ℓ_p norm of a vector $\boldsymbol{x} \in \mathbb{R}^n$. When $p \geq 1$, $\|\boldsymbol{x}\|_p$ is defined as:

$$\|\boldsymbol{x}\|_p \doteq \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}. \tag{1.3}$$

In particular, we care about the following three norms:

(i) $p = 1$, the ℓ_1 norm: $\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|$, *i.e.,* the sum of absolute values.

(ii) $p = 2$, the ℓ_2 norm (Euclidean norm): $\|\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$, *i.e.,* the length of vector $\boldsymbol{x}$.

(iii) $p = \infty$, the ℓ_∞ norm: $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$, *i.e.,* the maximum of absolute values.

For applications of ℓ_∞ norm in computer vision, one can refer to [6, 5] and the CVPR 2018 tutorial.[1]

Exercise: Verify the three norms ($\ell_1, \ell_2, \ell_\infty$) satisfy the properties in Definition 1.

Angle. The angle between two *nonzero* vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ is defined as:

$$\angle(\boldsymbol{x}, \boldsymbol{y}) = \arccos \left(\frac{\boldsymbol{x}^\top \boldsymbol{y}}{\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2} \right), \tag{1.4}$$

where we take $\arccos(\cdot) \in [0, \pi]$. We say $\boldsymbol{x}$ and $\boldsymbol{y}$ are *orthogonal* when $\boldsymbol{x}^\top \boldsymbol{y} = 0$. In machine learning, *cosine similarity*, *i.e.,* the cosine of the angle $\angle(\boldsymbol{x}, \boldsymbol{y})$, is often used to measure the similarity of two vectors $\boldsymbol{x}, \boldsymbol{y}$.

Frobenius Norm. The most common norm on $\mathbb{R}^{m \times n}$ is the *Frobenius norm*, For $\boldsymbol{X} \in \mathbb{R}^{m \times n}$, its Frobenius norm is defined as:

$$\|\boldsymbol{X}\|_F = \sqrt{\operatorname{tr}(\boldsymbol{X}^\top \boldsymbol{X})} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij}^2} = \|\operatorname{vec}(\boldsymbol{M})\|_2. \tag{1.5}$$

Operator Norm. Suppose $\|\cdot\|_p$ ($p \geq 1$) is a norm on $\mathbb{R}^n$ and $\mathbb{R}^m$, then we can define the *operator norm* (induced norm) of $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ as:

$$\|\boldsymbol{X}\|_p = \sup_{\boldsymbol{v} \in \mathbb{R}^n, \|\boldsymbol{v}\|_p \leq 1} \|\boldsymbol{X}\boldsymbol{v}\|_p. \tag{1.6}$$

A special case of the operator norm is $\|\boldsymbol{X}\|_2 = \sigma_{\max}(\boldsymbol{X})$, where $\sigma_{\max}(\boldsymbol{X})$ denotes the maximum singular value of $\boldsymbol{X}$. In general, $\|\boldsymbol{X}\|_p$ is NP-hard to compute for $p \notin \{1, 2, \infty\}$.

Exercise: Verify the matrix operator norm defined in eq. (1.6) satisfies Definition 1.

1.1.2 Trace, Vectorization and Kronecker Product

Cyclic Property. The trace operator is invariant under cyclic permutations:

$$\operatorname{tr}(\boldsymbol{ABCD}) = \operatorname{tr}(\boldsymbol{BCDA}) = \operatorname{tr}(\boldsymbol{CDAB}) = \operatorname{tr}(\boldsymbol{DABC}). \tag{1.7}$$

Moreover, if $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in \mathcal{S}^n$, then the trace operator is invariant under all permutations:

$$\operatorname{tr}(\boldsymbol{ABC}) = \operatorname{tr}\left((\boldsymbol{ABC})^\top \right) = \operatorname{tr}(\boldsymbol{CBA}) = \operatorname{tr}(\boldsymbol{ACB}). \tag{1.8}$$

[1] https://cs.adelaide.edu.au/~tjchin/tutorials/cvpr18/

Kronecker Product. If $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{p\times q}$, then the Kronecker product $\boldsymbol{A} \otimes \boldsymbol{B}$ is defined as:

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} A_{11}\boldsymbol{B} & \cdots & A_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ A_{m1}\boldsymbol{B} & \cdots & A_{mn}\boldsymbol{B} \end{bmatrix} \in \mathbb{R}^{mp\times nq}. \tag{1.9}$$

Useful Equalities. The following equalities can be useful when manipulating mathematical equations:

(i) If $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$, then $\operatorname{tr}\left(\boldsymbol{A}^\top \boldsymbol{B}\right) = \operatorname{vec}(\boldsymbol{A})^\top \operatorname{vec}(\boldsymbol{B})$.

(ii) Let $\boldsymbol{A} \in \mathbb{R}^{k\times l}, \boldsymbol{B} \in \mathbb{R}^{l\times m}, \boldsymbol{C} \in \mathbb{R}^{m\times n}$, then:

$$\operatorname{vec}(\boldsymbol{ABC}) = \left(\boldsymbol{C}^\top \otimes \boldsymbol{A}\right)\operatorname{vec}(\boldsymbol{B}) = (\mathbf{I}_n \otimes \boldsymbol{AB})\operatorname{vec}(\boldsymbol{C}) = \left(\boldsymbol{C}^\top\boldsymbol{B}^\top \otimes \mathbf{I}_k\right)\operatorname{vec}(\boldsymbol{A}), \tag{1.10}$$

$$\operatorname{vec}(\boldsymbol{AB}) = (\mathbf{I}_m \otimes \boldsymbol{A})\operatorname{vec}(\boldsymbol{B}) = \left(\boldsymbol{B}^\top \otimes \mathbf{I}_k\right)\operatorname{vec}(\boldsymbol{A}). \tag{1.11}$$

(iii) Let $\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{B}, \boldsymbol{Y}$ be real matrices with proper dimensions, then:

$$\operatorname{tr}\left(\boldsymbol{A}^\top\boldsymbol{X}^\top\boldsymbol{B}\boldsymbol{Y}\right) = \operatorname{vec}(\boldsymbol{X})^\top(\boldsymbol{A}\otimes\boldsymbol{B})\operatorname{vec}(\boldsymbol{Y}). \tag{1.12}$$

(iv) Let $\boldsymbol{X} \in \mathbb{R}^{m\times n}$, then its (i,j)-th entry can be written as:

$$X_{ij} = \boldsymbol{e}_i^\top\boldsymbol{X}\boldsymbol{e}_j = \operatorname{tr}\left(\boldsymbol{e}_i^\top\boldsymbol{X}\boldsymbol{e}_j\right) = \operatorname{tr}\left(\boldsymbol{X}\boldsymbol{e}_j\boldsymbol{e}_i^\top\right), \tag{1.13}$$

where $\boldsymbol{e}_i \in \mathbb{R}^m$ is the i-th standard basis vector (1 at the i-th entry and 0 everywhere else), and $\boldsymbol{e}_j \in \mathbb{R}^n$ is the j-th standard basis vector (1 at the j-th entry and 0 everywhere else).

The interested reader can refer to the supplementary material of [4] for an application of the equalities above to solving a problem in computer vision.

1.1.3 Orthogonal Matrices

An *orthogonal matrix* is a real *square* matrix whose rows and columns are orthonormal vectors (orthogonal and unit norm). Formally, let $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$, then $\boldsymbol{Q}$ is an orthogonal matrix if and only if:

$$\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^\top = \mathbf{I}_n \Longleftrightarrow \boldsymbol{Q}^{-1} = \boldsymbol{Q}^\top. \tag{1.14}$$

We use $\mathrm{O}(n)$, the n-dimensional *orthogonal group*, to denote the set of orthogonal matrices with size $n \times n$. An orthogonal matrix has determinant equal to either $+1$ or -1, which can be easily seen from:

$$\det\left(\boldsymbol{Q}^\top\boldsymbol{Q}\right) = (\det(\boldsymbol{Q}))^2 = \det(\mathbf{I}_n) = 1 \Longrightarrow \det(\boldsymbol{Q}) = \pm 1. \tag{1.15}$$

An orthogonal matrix preserves inner product in Euclidean space:

$$\langle \boldsymbol{Qx}, \boldsymbol{Qy}\rangle = (\boldsymbol{Qx})^\top\boldsymbol{Qy} = \boldsymbol{x}^\top\boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{y} = \boldsymbol{x}^\top\boldsymbol{y} = \langle\boldsymbol{x},\boldsymbol{y}\rangle, \quad \forall \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n, \boldsymbol{Q} \in \mathrm{O}(n). \tag{1.16}$$

As a result, orthogonal matrix is also ℓ_2-norm preserving (*cf.* eq. (1.1)):

$$\|\boldsymbol{Qx}\|_2 = \|\boldsymbol{x}\|_2, \quad \forall \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{Q} \in \mathrm{O}(n). \tag{1.17}$$

Special Orthogonal Matrices. A special orthogonal matrix is an orthogonal matrix whose determinant is equal to $+1$. We use $\mathrm{SO}(n)$, the n-dimensional *special orthogonal group*, to denote the set of special orthogonal matrices with size $n \times n$, *i.e.,* $\mathrm{SO}(n) \doteq \{\boldsymbol{R} \in \mathbb{R}^{n\times n} : \boldsymbol{R}^\mathsf{T}\boldsymbol{R} = \boldsymbol{R}\boldsymbol{R}^\mathsf{T} = \mathbf{I}_n, \det(\boldsymbol{R}) = +1\}$. Moreover, $\mathrm{SO}(n)$ is the set of proper *rotation* matrices in Euclidean space. For example, $\mathrm{SO}(3)$ describes the set of proper 3D rotations, which is ubiquitous in robotics and computer vision applications.

Projection onto $\mathrm{O}(n)$ and $\mathrm{SO}(n)$. Given an arbitrary square matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$, projecting $\boldsymbol{M}$ onto $\mathrm{O}(n)$ (resp. $\mathrm{SO}(n)$) consists in finding the matrix $\boldsymbol{Q} \in \mathrm{O}(n)$ (resp. $\boldsymbol{R} \in \mathrm{SO}(n)$) that is *nearest* to $\boldsymbol{M}$:

$$\Pi_{\mathrm{O}(n)}(\boldsymbol{M}) = \operatorname*{arg\,min}_{\boldsymbol{Q}\in\mathrm{O}(n)} \|\boldsymbol{Q}-\boldsymbol{M}\|_F^2\,; \quad \Pi_{\mathrm{SO}(n)}(\boldsymbol{M}) = \operatorname*{arg\,min}_{\boldsymbol{R}\in\mathrm{SO}(n)} \|\boldsymbol{R}-\boldsymbol{M}\|_F^2\,. \tag{1.18}$$

This is a notable example of tractable nonconvex problems. In fact, the two projection problems in eq. (1.18) admit closed-form solutions, using *singular value decomposition* (SVD) (*cf.* Section 1.1.4) [7, 1]. Formally, let $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\mathsf{T}$ be the SVD of $\boldsymbol{M}$, where $\boldsymbol{U}, \boldsymbol{V} \in \mathrm{O}(n)$, and $\boldsymbol{S}$ contains the singular values of $\boldsymbol{M}$ in *descending* order,[2] *i.e.,* $S_{11} \geq S_{22} \geq \ldots \geq S_{nn} \geq 0$, then:

$$\Pi_{\mathrm{O}(n)}(\boldsymbol{M}) = \boldsymbol{U}\boldsymbol{V}^\mathsf{T}, \tag{1.19}$$

$$\Pi_{\mathrm{SO}(n)}(\boldsymbol{M}) = \boldsymbol{U}\operatorname{diag}([1,1,\ldots,\det(\boldsymbol{U})\det(\boldsymbol{V})])\,\boldsymbol{V}^\mathsf{T}. \tag{1.20}$$

Proof. We only prove the solution for projection onto $\mathrm{O}(n)$ (eq. (1.19)), while we leave the proof for projection onto $\mathrm{SO}(n)$ as an exercise. To prove eq. (1.19) is the solution for problem (1.18), we develop the cost function of problem (1.18):

$$\operatorname*{arg\,min}_{\boldsymbol{Q}\in\mathrm{O}(n)} \|\boldsymbol{Q}-\boldsymbol{M}\|_F^2 = \operatorname*{arg\,min}_{\boldsymbol{Q}\in\mathrm{O}(n)} \operatorname{tr}\left((\boldsymbol{Q}-\boldsymbol{M})^\mathsf{T}(\boldsymbol{Q}-\boldsymbol{M})\right) \tag{1.21}$$

$$= \operatorname*{arg\,min}_{\boldsymbol{Q}\in\mathrm{O}(n)} \operatorname{tr}\left(\boldsymbol{Q}^\mathsf{T}\boldsymbol{Q}\right) + \operatorname{tr}\left(\boldsymbol{M}^\mathsf{T}\boldsymbol{M}\right) - 2\operatorname{tr}\left(\boldsymbol{Q}^\mathsf{T}\boldsymbol{M}\right) = \operatorname*{arg\,max}_{\boldsymbol{Q}\in\mathrm{O}(n)} \operatorname{tr}\left(\boldsymbol{Q}^\mathsf{T}\boldsymbol{M}\right) \tag{1.22}$$

$$= \operatorname*{arg\,max}_{\boldsymbol{Q}\in\mathrm{O}(n)} \operatorname{tr}\left(\boldsymbol{Q}^\mathsf{T}\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\mathsf{T}\right) = \operatorname*{arg\,max}_{\boldsymbol{Q}\in\mathrm{O}(n)} \operatorname{tr}\left(\boldsymbol{S}\boldsymbol{V}^\mathsf{T}\boldsymbol{Q}^\mathsf{T}\boldsymbol{U}\right), \tag{1.23}$$

and observe that $\boldsymbol{V}^\mathsf{T}\boldsymbol{Q}^\mathsf{T}\boldsymbol{U}$ is an orthogonal matrix because $\boldsymbol{V}, \boldsymbol{Q}, \boldsymbol{U} \in \mathrm{O}(n)$. Therefore, as $\boldsymbol{S}$ is an diagonal matrix, the maximum of $\operatorname{tr}\left(\boldsymbol{S}\boldsymbol{V}^\mathsf{T}\boldsymbol{Q}^\mathsf{T}\boldsymbol{U}\right)$ is attained if and only if $\boldsymbol{V}^\mathsf{T}\boldsymbol{Q}^\mathsf{T}\boldsymbol{U} = \mathbf{I}_n$, *i.e.,* $\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{V}^\mathsf{T}$. □

1.1.4 Singular Value Decomposition

The *singular value decomposition* (SVD) of any real matrix[3] $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ is:

$$\boldsymbol{M} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\mathsf{T}, \quad \boldsymbol{U} \in \mathrm{O}(m), \boldsymbol{V} \in \mathrm{O}(n), \tag{1.24}$$

and $\boldsymbol{S} \in \mathbb{R}^{m\times n}$ is a rectangular diagonal matrix with nonnegative diagonal entries. The diagonal entries of $\boldsymbol{S}$ are called the *singular values* of $\boldsymbol{M}$. The number of nonzero singular values in $\boldsymbol{S}$ is equal to the rank of $\boldsymbol{M}$. The SVD in eq. (1.24) in equivalent to:

$$\boldsymbol{M}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{S}, \quad \boldsymbol{U} \in \mathrm{O}(m), \boldsymbol{V} \in \mathrm{O}(n), \tag{1.25}$$

which implies that $\boldsymbol{M}\boldsymbol{v}_i = S_{ii}\boldsymbol{u}_i, i = 1,\ldots,\min\{m,n\}$, where $\boldsymbol{u}_i \in \mathbb{R}^m$, $\boldsymbol{v}_i \in \mathbb{R}^n$ are the i-th column of $\boldsymbol{U}$ and $\boldsymbol{V}$, and they are called the left and right *singular vectors* of $\boldsymbol{M}$, respectively.

Exercise: Is the SVD of a matrix unique? What is an SVD of an orthogonal matrix and a rotation matrix?

---

[2] For example, Matlab `svd` returns the singular values in descending order.

[3] Complex matrices also admit SVD factorizations.

Relationship to Spectral Decomposition. Consider matrices $\boldsymbol{M}^\top\boldsymbol{M}$ and $\boldsymbol{M}\boldsymbol{M}^\top$:

$$\boldsymbol{M}^\top\boldsymbol{M} = \boldsymbol{V}\boldsymbol{S}^\top\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top = \boldsymbol{V}\left(\boldsymbol{S}^\top\boldsymbol{S}\right)\boldsymbol{V}^\top, \tag{1.26}$$

$$\boldsymbol{M}\boldsymbol{M}^\top = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{S}^\top\boldsymbol{U}^\top = \boldsymbol{U}\left(\boldsymbol{S}\boldsymbol{S}^\top\right)\boldsymbol{U}^\top. \tag{1.27}$$

Therefore, the columns of $\boldsymbol{V}$ are eigenvectors of $\boldsymbol{M}^\top\boldsymbol{M}$, while the columns of $\boldsymbol{U}$ are eigenvectors of $\boldsymbol{M}\boldsymbol{M}^\top$. The nonzero singular values of $\boldsymbol{M}$ are the square roots of the nonzero eigenvalues of $\boldsymbol{M}^\top\boldsymbol{M}$ and $\boldsymbol{M}\boldsymbol{M}^\top$.

1.1.5 Eigenvalues and Eigenvectors

Given a square matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, if there exists a scalar $\lambda \in \mathbb{C}$ and a nonzero vector $\boldsymbol{v} \in \mathbb{C}^n$ such that:

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}, \tag{1.28}$$

then $\boldsymbol{v}$ is called a *right eigenvector* of $\boldsymbol{A}$ and λ is the associated eigenvalue. If there exists a scalar $\kappa \in \mathbb{C}$ and a nonzero vector $\boldsymbol{u} \in \mathbb{C}^n$ such that:

$$\boldsymbol{u}^\top\boldsymbol{A} = \kappa\boldsymbol{u}^\top, \tag{1.29}$$

then $\boldsymbol{u}$ is called a *left eigenvector* of $\boldsymbol{A}$ with associated eigenvalue κ. If all right eigenvectors of $\boldsymbol{A}$ are *linearly independent*, then denoting $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n]$ and $\boldsymbol{\Lambda} = \operatorname{diag}\left([\lambda_1, \ldots, \lambda_n]\right)$, we have:

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Lambda} \Longrightarrow \boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V} = \boldsymbol{\Lambda}, \tag{1.30}$$

i.e., $\boldsymbol{A}$ is *diagonalizable* by $\boldsymbol{V}$.

As an exercise, try to prove the following lemmas.

Lemma 2 (Eigenvalues and Characteristic Polynomial)**.** *Any matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ has* equal *left and right eigenvalues, and they are the roots of the characteristic polynomial $f(\lambda) = det(\boldsymbol{A} - \lambda\mathbf{I}_n)$.*

Lemma 3 (Real Symmetric Matrices have Real Eigenvalues)**.** *The eigenvalues of any real symmetric matrix $\boldsymbol{A} \in \mathcal{S}^n$ are* all real. *Hence, the eigenvalues can be sorted: $\lambda_1 \geq \ldots \geq \lambda_n$.*

Lemma 4 (Real Symmetric Matrices have Orthogonal Eigenvectors)**.** *Let $\boldsymbol{A} \in \mathcal{S}^n$ be a real symmetric matrix and let $\lambda_i \neq \lambda_j$ be any two distinct eigenvalues with associated eigenvectors $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$, then $\boldsymbol{v}_i^\top\boldsymbol{v}_j = 0$. Moreover, if λ_i is a repeated eigenvalue with multiplicity $m \geq 2$, then there exist m orthonormal eigenvectors corresponding to λ_i.*

Corollary 5 (Real Symmetric Matrices are Diagonalizable)**.** *Any real symmetric matrix $\boldsymbol{A} \in \mathcal{S}^n$ can be diagonalized as:*

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top, \tag{1.31}$$

where $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n] \in \mathrm{O}(n)$ is an orthogonal matrix whose columns $\boldsymbol{u}_i$ are the eigenvectors of $\boldsymbol{A}$, and $\boldsymbol{\Lambda} = \operatorname{diag}\left([\lambda_1, \ldots, \lambda_n]\right)$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{A}$. The factorization in eq. (1.31) is called the eigendecomposition or spectral decomposition of $\boldsymbol{A}$, and is unique (up to permutation of $\boldsymbol{u}_i$ and λ_i) when all the eigenvalues of $\boldsymbol{A}$ are distinct.

Some useful properties of eigenvalues and eigenvectors:

(i) Trace of a matrix equals the sum of all eigenvalues: $\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^n \lambda_i$, for all $\boldsymbol{A} \in \mathbb{C}^{n\times n}$.

(ii) Determinant of a matrix equals the product of all eigenvalues: $\det(\boldsymbol{A}) = \prod_{i=1}^n \lambda_i$, for all $\boldsymbol{A} \in \mathbb{C}^{n\times n}$.

(iii) The eigenvalues of $\boldsymbol{A}^k$, *i.e.,* the k-th power of $\boldsymbol{A}$, is $\lambda_i^k, i = 1, \ldots, n$, for any $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ and $k \in \mathbb{N}$. More generally, let $f(\cdot)$ be any polynomial, then the eigenvalues of $f(\boldsymbol{A})$ are $f(\lambda_i), i = 1, \ldots, n$.

(iv) The eigenvalues of $\boldsymbol{A}^{-1}$ are $\lambda_i^{-1}, i = 1, \ldots, n$, for any $\boldsymbol{A} \in \mathbb{C}^{n\times n}$.

(v) The eigenvalues of $\boldsymbol{A} + \alpha\mathbf{I}_n$ are $\lambda_i + \alpha, i = 1, \ldots, n$, for any $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ and $\alpha \in \mathbb{C}$.

1.1.6 Positive Semidefinite Matrices

The following statements about *positive semidefinite* (PSD) matrices are equivalent:

(i) The matrix $\boldsymbol{A} \in \mathcal{S}^n$ is positive semidefinite ($\boldsymbol{A} \succeq 0$, $\boldsymbol{A} \in \mathcal{S}^n_+$).

(ii) For all $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x}^\mathsf{T} \boldsymbol{A} \boldsymbol{x} \geq 0$.

(iii) All eigenvalues of $\boldsymbol{A}$ are nonnegative.

(iv) All $2^n - 1$ principal minors of $\boldsymbol{A}$ are nonnegative.

(v) There exists a factorization $\boldsymbol{A} = \boldsymbol{B}^\mathsf{T} \boldsymbol{B}$.

The following statements about *positive definite* (PD) matrices are equivalent:

(i) The matrix $\boldsymbol{A} \in \mathcal{S}^n$ is positive definite ($\boldsymbol{A} \succ 0$, $\boldsymbol{A} \in \mathcal{S}^n_{++}$).

(ii) For all *nonzero* $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x}^\mathsf{T} \boldsymbol{A} \boldsymbol{x} > 0$.

(iii) All eigenvalues of $\boldsymbol{A}$ are strictly positive.

(iv) All n leading principal minors of $\boldsymbol{A}$ are positive.

(v) There exists a factorization $\boldsymbol{A} = \boldsymbol{B}^\mathsf{T} \boldsymbol{B}$, with $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ and $\operatorname{rank}(\boldsymbol{B}) = n$.

Matrix Congruence. If $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ is invertible (nonsingular), then $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$ are congruent if:

$$\boldsymbol{P}^\mathsf{T} \boldsymbol{A} \boldsymbol{P} = \boldsymbol{B}. \tag{1.32}$$

If both $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric, then $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same numbers of positive, negative, and zero eigenvalues. Therefore, $\boldsymbol{A} \succeq 0 \Longleftrightarrow \boldsymbol{P}^\mathsf{T} \boldsymbol{A} \boldsymbol{P} \succeq 0$.

Matrix Similarity. If $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ is invertible (nonsingular), then $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$ are similar if:

$$\boldsymbol{P}^{-1} \boldsymbol{A} \boldsymbol{P} = \boldsymbol{B}. \tag{1.33}$$

Similar matrices have the same characteristic polynomials, hence, the same eigenvalues. (*Exercise:* prove this.) Therefore, $\boldsymbol{A} \succeq 0 \Longleftrightarrow \boldsymbol{P}^{-1} \boldsymbol{A} \boldsymbol{P} \succeq 0$. See [8] for an application of this.

Schur Complement. Consider the following block matrix:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} \in \mathbb{R}^{(m+n)\times(m+n)}, \quad \boldsymbol{A} \in \mathbb{R}^{m \times m}, \boldsymbol{B} \in \mathbb{R}^{m \times n}, \boldsymbol{C} \in \mathbb{R}^{n \times m}, \boldsymbol{D} \in \mathbb{R}^{n \times n}. \tag{1.34}$$

If block $\boldsymbol{A}$ is invertible, then the *Schur complement* of block $\boldsymbol{A}$ of $\boldsymbol{X}$ is:

$$\boldsymbol{X}/\boldsymbol{A} \doteq \boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}. \tag{1.35}$$

If block $\boldsymbol{D}$ is invertible, then the *Schur complement* of block $\boldsymbol{D}$ of $\boldsymbol{X}$ is:

$$\boldsymbol{X}/\boldsymbol{D} \doteq \boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}. \tag{1.36}$$

In the case that $\boldsymbol{A}$ or $\boldsymbol{D}$ is singular, replacing $\boldsymbol{A}^{-1}$ and $\boldsymbol{D}^{-1}$ with *generalized inverses*[4] yields the *generalized Schur complement*.

[4] https://en.wikipedia.org/wiki/Generalized_inverse

Schur complement is one of the most important tools for analyzing positive semi-definiteness and positive definiteness of symmetric matrices. Consider the following symmetric matrix:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^{\mathsf{T}} & \boldsymbol{D} \end{bmatrix} \in \mathcal{S}^{(m+n)}, \quad \boldsymbol{A} \in \mathbb{R}^{m\times m}, \boldsymbol{B} \in \mathbb{R}^{m\times n}, \boldsymbol{D} \in \mathbb{R}^{n\times n}, \tag{1.37}$$

then we have:

(i) Sufficient and necessary conditions for $\boldsymbol{X} \succ 0$:

$$\boldsymbol{X} \succ 0 \quad \Longleftrightarrow \quad \boldsymbol{A} \succ 0, \boldsymbol{X}/\boldsymbol{A} \succ 0 \quad \Longleftrightarrow \quad \boldsymbol{D} \succ 0, \boldsymbol{X}/\boldsymbol{D} \succ 0. \tag{1.38}$$

(ii) Sufficient and necessary condition for $\boldsymbol{X} \succeq 0$ when $\boldsymbol{A}$ is invertible:

$$\boldsymbol{X} \succeq 0 \quad \Longleftrightarrow \quad \boldsymbol{A} \succ 0, \boldsymbol{X}/\boldsymbol{A} \succeq 0. \tag{1.39}$$

(iii) Sufficient and necessary condition for $\boldsymbol{X} \succeq 0$ when $\boldsymbol{D}$ is invertible:

$$\boldsymbol{X} \succeq 0 \quad \Longleftrightarrow \quad \boldsymbol{D} \succ 0, \boldsymbol{X}/\boldsymbol{D} \succeq 0. \tag{1.40}$$

(iv) Sufficient and necessary condition for $\boldsymbol{X} \succeq 0$ can be described using generalized Schur complements [9].

1.2 Matrix Calculus

A good online resource for matrix calculus: `http://www.matrixcalculus.org/`.

A good book for matrix calculus: The Matrix Cookbook.

1.2.1 Derivative and Gradient

Let $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ be a differentiable function:

$$\boldsymbol{f}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})]^{\mathsf{T}}, \tag{1.41}$$

where each $f_i, i = 1, \ldots, m$ is differentiable. Then the *Jacobian* (or derivative) of $\boldsymbol{f}$ w.r.t. $\boldsymbol{x}$, denoted as $D\boldsymbol{f}(\boldsymbol{x})$, is an $m \times n$ matrix whose (i, j)-th entry is:

$$[D\boldsymbol{f}(\boldsymbol{x})]_{ij} = \frac{\partial f_i(\boldsymbol{x})}{\partial x_j}, i = 1, \ldots, m; j = 1, \ldots, n. \tag{1.42}$$

In other word, the i-th row of the Jacobian is the derivative of f_i w.r.t. $\boldsymbol{x}$:

$$D\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} Df_1(\boldsymbol{x}) \\ \vdots \\ Df_m(\boldsymbol{x}) \end{bmatrix} \in \mathbb{R}^{m\times n}, \quad Df_i(\boldsymbol{x}) \in \mathbb{R}^{1\times n}, i = 1, \ldots, m. \tag{1.43}$$

In the case when f is a real-valued function, *i.e.,* $f : \mathbb{R}^n \to \mathbb{R}$ (*e.g.,* each f_i in eq. (1.41)), then the *gradient* of f w.r.t. $\boldsymbol{x}$ is the transpose of $Df(\boldsymbol{x})$:

$$\nabla f(\boldsymbol{x}) = Df(\boldsymbol{x})^{\mathsf{T}} \in \mathbb{R}^n, \tag{1.44}$$

which is a column vector.

We can use the Jacobian to perform first-order approximation of a function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ at $\boldsymbol{x}$:

$$\boldsymbol{f}(\boldsymbol{z}) \approx \boldsymbol{f}(\boldsymbol{x}) + D\boldsymbol{f}(\boldsymbol{x}) \cdot (\boldsymbol{z} - \boldsymbol{x}). \tag{1.45}$$

Chain Rule. Suppose $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ and $\boldsymbol{g} : \mathbb{R}^m \to \mathbb{R}^p$ are both differentiable, then the composition $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^p$ defined by $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}))$ is differentiable at $\boldsymbol{x}$, with the derivative computed by the chain rule as:

$$D\boldsymbol{h}(\boldsymbol{x}) = D\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}))\, D\boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^{p\times n}, \quad D\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x})) \in \mathbb{R}^{p\times m}, D\boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^{m\times n}. \tag{1.46}$$

1.2.2 Second Derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function, then the second derivative, *i.e.,* the *Hessian*, of f w.r.t. $\boldsymbol{x}$, denoted as $\nabla^2 f(\boldsymbol{x})$, is:

$$\left[\nabla^2 f(\boldsymbol{x})\right]_{ij} = \frac{\partial f(\boldsymbol{x})}{\partial x_i \partial x_j}, i = 1, \ldots, n; j = 1, \ldots, n. \tag{1.47}$$

By definition, the Hessian $\nabla^2 f(\boldsymbol{x}) \in \mathcal{S}^n$ is a symmetric matrix. The Hessian can be interpreted as the derivative of the gradient: $\nabla^2 f(\boldsymbol{x}) = D\nabla f(\boldsymbol{x})$.

Using the gradient and Hessian of f, the second-order approximation of f at $\boldsymbol{x}$ can be written as:

$$f(\boldsymbol{z}) \approx f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\mathsf{T}(\boldsymbol{z} - \boldsymbol{x}) + \frac{1}{2}(\boldsymbol{z} - \boldsymbol{x})^\mathsf{T} \nabla^2 f(\boldsymbol{x})(\boldsymbol{z} - \boldsymbol{x}). \tag{1.48}$$

Exercise: Let $f : \mathbb{R}^n \to \mathbb{R}$, and $g : \mathbb{R} \to \mathbb{R}$ be two twice differentiable functions, what is the Hessian of $h(\boldsymbol{x}) = g(f(\boldsymbol{x}))$?

References

[1] K.S. Arun, T.S. Huang, and S.D. Blostein. Least-squares fitting of two 3-D point sets. *IEEE Trans. Pattern Anal. Machine Intell.*, 9(5):698–700, sept. 1987.

[2] Grigoriy Blekherman, Pablo A Parrilo, and Rekha R Thomas. *Semidefinite optimization and convex algebraic geometry.* SIAM, 2012.

[3] S. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge University Press, 2004.

[4] Jesus Briales, Laurent Kneip, and Javier Gonzalez-Jimenez. A certifiably globally optimal solution to the non-minimal relative pose problem. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[5] Olof Enqvist and Fredrik Kahl. Robust optimal pose estimation. In *European Conf. on Computer Vision (ECCV)*, pages 141–153. Springer, 2008.

[6] F. Kahl and R. Hartley. Multiple-view geometry under the ℓ_∞-norm. *IEEE Trans. Pattern Anal. Machine Intell.*, 30(9):1603–1617, 2008.

[7] P.H. Schonemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31:1–10, 1966.

[8] H. Yang and L. Carlone. A quaternion-based certifiably optimal solution to the Wahba problem with outliers. In *Intl. Conf. on Computer Vision (ICCV)*, 2019.

[9] Fuzhen Zhang. *The Schur complement and its applications*, volume 4. Springer Science & Business Media, 2006.