# EE/ACM 150 - Applications of Convex Optimization in Signal Processing and Communications
# Lecture 2

Andre Tkacenko

Signal Processing Research Group
Jet Propulsion Laboratory

April 5, 2012

Caltech

# Outline

1. Linear Algebra / Matrix Analysis Notation & Definitions
2. Basic Vector Space Results
3. Overview of Matrix Analysis Concepts
4. Special Types of Matrices
5. Inner Products & Norms

# Terminology List

- Common fields and sets

| | | |
|---|---|---|
| $\mathbb{R}$ | – | field of real scalars |
| $\mathbb{C}$ | – | field of complex scalars |
| $\mathbb{F}$ | – | general field (will be either $\mathbb{R}$ or $\mathbb{C}$ here) |
| $\mathbb{R}^n, \mathbb{C}^n, \mathbb{F}^n$ | – | set of $n \times 1$ vectors over $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{F}$, respectively |
| $\mathbb{R}^{m \times n}, \mathbb{C}^{m \times n}, \mathbb{F}^{m \times n}$ | – | set of $m \times n$ matrices over $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{F}$, respectively |

- Special vector/matrix sets

| | | |
|---|---|---|
| $\mathbb{R}^n_+, \mathbb{R}^n_{++}$ | – | set of $n \times 1$ real vectors whose components are nonnegative or positive, respectively |
| $\mathbb{S}^n$ | – | set of $n \times n$ real symmetric matrices |
| $\mathbb{H}^n$ | – | set of $n \times n$ Hermitian matrices |
| $\mathbb{S}^n_+, \mathbb{S}^n_{++}$ | – | set of $n \times n$ real symmetric positive semidefinite or positive definite matrices, respectively |
| $\mathbb{H}^n_+, \mathbb{H}^n_{++}$ | – | set of $n \times n$ Hermitian positive semidefinite or positive definite matrices, respectively |

# Terminology List (Continued)

- Vector space quantities

$\dim(\mathcal{V})$ – dimension of vector space $\mathcal{V}$

$\mathcal{R}(\mathbf{A})$ – range space of matrix $\mathbf{A}$

$\mathcal{N}(\mathbf{A})$ – null space of matrix $\mathbf{A}$

$\mathrm{rank}(\mathbf{A})$ – rank of matrix $\mathbf{A}$ (i.e., $\dim(\mathcal{R}(\mathbf{A}))$)

$\mathrm{nullity}(\mathbf{A})$ – nullity of matrix $\mathbf{A}$ (i.e., $\dim(\mathcal{N}(\mathbf{A}))$)

- Common matrix operators and quantities

$*$ – complex conjugate operator (i.e., $\mathbf{a}^*$ or $\mathbf{A}^*$)

$T$ – transpose operator (i.e., $\mathbf{a}^T$ or $\mathbf{A}^T$)

$\dagger$ – complex conjugate transpose operator (i.e., $\mathbf{a}^\dagger$ or $\mathbf{A}^\dagger$)

$\det(\mathbf{A})$ – determinant of *square* matrix $\mathbf{A}$

$\mathbf{A}^{-1}$ – inverse of *square* matrix $\mathbf{A}$ (if it exists)

$\mathrm{diag}(\mathbf{A})$ – column vector formed from diagonal components of matrix $\mathbf{A}$

$\mathrm{diag}(\mathbf{a})$ – diagonal matrix formed from components of vector $\mathbf{a}$

$\mathrm{tr}(\mathbf{A})$ – trace of matrix $\mathbf{A}$ (i.e., the sum of diagonal components)

# Terminology List (Continued)

- Inner product / norm quantities

| | | |
|---|---|---|
| $\langle \mathbf{x}, \mathbf{y} \rangle$ | – | inner product of $\mathbf{x}$ and $\mathbf{y}$ ($\mathbf{x}$ and $\mathbf{y}$ can be either vectors or matrices) |
| $\|\|\mathbf{a}\|\|, \|\|\mathbf{A}\|\|$ | – | norm of vector $\mathbf{a}$ or matrix $\mathbf{A}$, respectively |
| $\|\|\mathbf{a}\|\|_p$ | – | $\ell_p$-norm of vector $\mathbf{a}$ |
| $\|\|\mathbf{A}\|\|_{\mathrm{a,b}}$ | – | operator norm of $\mathbf{A}$ induced by vector norms $\|\|\cdot\|\|_{\mathrm{a}}$ and $\|\|\cdot\|\|_{\mathrm{b}}$ |
| $\|\|\mathbf{A}\|\|_F$ | – | Frobenius norm of matrix $\mathbf{A}$ |
| $\|\|\mathbf{z}\|\|_{\circledast}$ | – | dual norm of $\mathbf{z}$ associated with $\|\|\mathbf{z}\|\|$ ($\mathbf{z}$ can be either a vector or a matrix) |

- Useful vectors & matrices and miscellaneous terminology

| | | |
|---|---|---|
| $\mathbf{1}$ | – | column vector of all ones |
| $\mathbf{0}_{m \times n}$ | – | $m \times n$ matrix of zeros |
| $\mathbf{I}_n$ | – | $n \times n$ identity matrix |
| $[\mathbf{a}]_k$ | – | $k$-th element of vector $\mathbf{a}$ |
| $[\mathbf{A}]_{k,\ell}$ | – | $(k, \ell)$-th element of matrix $\mathbf{A}$ |

# Linear Dependence and Independence

A set of $n$ vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ from a vector space $\mathcal{V}$ is said to be *linearly dependent* if and only if there are $n$ scalars $a_1, a_2, \ldots, a_n$ *not all zero* such that

$$\sum_{k=1}^{n} a_k \mathbf{v}_k = \mathbf{0}\,.$$

In other words, there is at least one vector, say $\mathbf{v}_\ell$, which depends linearly on the other vectors, i.e.,

$$\mathbf{v}_\ell = -\frac{1}{a_\ell} \sum_{\substack{k=1 \\ k\neq\ell}}^{n} a_k \mathbf{v}_k\,.$$

The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are said to be *linearly independent* if they are not linearly dependent. Equivalently, the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly independent when

$$\sum_{k=1}^{n} a_k \mathbf{v}_k = \mathbf{0}\,,$$

if and only if $a_1 = a_2 = \cdots = a_n = 0$.

# Span, Basis, & Dimension

Let $S \triangleq \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ denote a subset of vectors from a vector space $\mathcal{V}$ defined over $\mathbb{F}$. The *span* of $S$ (denoted $\mathrm{span}(S)$) is the set of all linear combinations of elements in $S$, i.e.,

$$\mathrm{span}(S) \triangleq \{c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n : c_1, \ldots, c_n \in \mathbb{F}\} \ .$$

If $S$ is a linearly independent set of vectors which spans $\mathcal{V}$ (meaning $\mathrm{span}(S) = \mathcal{V}$), then $S$ is said to be a *basis* for $\mathcal{V}$.

Any vector $\mathbf{w} \in \mathcal{V}$ can be represented in terms of a basis $S$ in one and only one way. As such, a basis effectively defines a "coordinate system" for $\mathcal{V}$.

A vector space $\mathcal{V}$ can be characterized by several different bases (bases are nonunique). However, all bases for a given vector space will have the same number of elements. This common number is call the *dimension* of the vector space $\mathcal{V}$ and is denoted $\dim(\mathcal{V})$.

# Range & Null Space, Rank & Nullity

Let $\mathbf{A} \in \mathbb{F}^{m \times n}$. Then, the range space $\mathcal{R}(\mathbf{A})$ and null space $\mathcal{N}(\mathbf{A})$ are defined as follows.

$$\begin{aligned} \mathcal{R}(\mathbf{A}) &\triangleq \{\mathbf{y} = \mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{F}^n\} \\ \mathcal{N}(\mathbf{A}) &\triangleq \{\mathbf{x} \in \mathbb{F}^n : \mathbf{A}\mathbf{x} = \mathbf{0}\} \end{aligned} .$$

The rank and nullity of $\mathbf{A}$ are the dimensions of the range and null spaces, respectively, i.e.,

$$\begin{aligned} \mathrm{rank}(\mathbf{A}) &\triangleq \dim(\mathcal{R}(\mathbf{A})) \\ \mathrm{nullity}(\mathbf{A}) &\triangleq \dim(\mathcal{N}(\mathbf{A})) \end{aligned} .$$

## Equality of row and column ranks:

$$\mathrm{rank}(\mathbf{A}) = \mathrm{rank}\left(\mathbf{A}^T\right) .$$

## Rank-nullity theorem:

$$\mathrm{rank}(\mathbf{A}) + \mathrm{nullity}(\mathbf{A}) = n .$$

# Systems of Linear Equations

Let $\mathbf{A} \in \mathbb{F}^{m \times n}$, $\mathbf{x} \in \mathbb{F}^{n}$, and $\mathbf{b} \in \mathbb{F}^{m}$. Consider the following system of linear equations which we would like to solve for $\mathbf{x}$.

$$\mathbf{A}\mathbf{x} = \mathbf{b} \, .$$

This system can behave in any one of three possible ways:

1. It can have a single unique solution. ($\mathbf{b} \in \mathcal{R}(\mathbf{A})$, $\text{nullity}(\mathbf{A}) = 0$)
2. It can have infinitely many solutions. ($\mathbf{b} \in \mathcal{R}(\mathbf{A})$, $\text{nullity}(\mathbf{A}) > 0$)
3. It can have no solution. ($\mathbf{b} \notin \mathcal{R}(\mathbf{A})$)

If a solution exists (i.e., $\mathbf{b} \in \mathcal{R}(\mathbf{A})$), then the set of solutions can be characterized as follows. Let $\mathbf{p}$ be a particular solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$, i.e., $\mathbf{A}\mathbf{p} = \mathbf{b}$. Then, the solution set is given by

$$\{\mathbf{p} + \mathbf{v} : \mathbf{A}\mathbf{v} = \mathbf{0}\} \; ,$$

that is, the solution set is a *translation* of the solution set of the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$.

# Matrix Arithmetic Operations

- **Addition and scalar multiplication:**
If $\alpha, \beta \in \mathbb{F}$, $\mathbf{A} \in \mathbb{F}^{m \times n}$, and $\mathbf{B} \in \mathbb{F}^{m \times n}$, then $\mathbf{C} \triangleq \alpha \mathbf{A} + \beta \mathbf{B}$ is such that $\mathbf{C} \in \mathbb{F}^{m \times n}$ with $[\mathbf{C}]_{k,\ell} = \alpha [\mathbf{A}]_{k,\ell} + \beta [\mathbf{B}]_{k,\ell}$.
- **Matrix multiplication:**
If $\mathbf{A} \in \mathbb{F}^{m \times n}$ and $\mathbf{B} \in \mathbb{F}^{n \times p}$, then $\mathbf{C} \triangleq \mathbf{AB}$ is such that $\mathbf{C} \in \mathbb{F}^{m \times p}$ with
$$[\mathbf{C}]_{k,\ell} = \sum_{i=1}^{n} [\mathbf{A}]_{k,i} [\mathbf{B}]_{i,\ell} \, , 1 \leq k \leq m, 1 \leq \ell \leq p \, .$$
- **Block matrix multiplication:**
If $\mathbf{A} \in \mathbb{F}^{m_1 \times n_1}$, $\mathbf{B} \in \mathbb{F}^{m_1 \times n_2}$, $\mathbf{C} \in \mathbb{F}^{m_2 \times n_1}$, $\mathbf{D} \in \mathbb{F}^{m_2 \times n_2}$, $\mathbf{E} \in \mathbb{F}^{n_1 \times p_1}$, $\mathbf{F} \in \mathbb{F}^{n_1 \times p_2}$, $\mathbf{G} \in \mathbb{F}^{n_2 \times p_1}$, and $\mathbf{H} \in \mathbb{F}^{n_2 \times p_2}$, then we have
$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix} = \begin{bmatrix} \mathbf{AE} + \mathbf{BG} & \mathbf{AF} + \mathbf{BH} \\ \mathbf{CE} + \mathbf{DG} & \mathbf{CF} + \mathbf{DH} \end{bmatrix} .$$

# Determinant of a Square Matrix

The determinant of an $n \times n$ square matrix $\mathbf{A}$, denoted $\det(\mathbf{A})$, is a scalar quantity used to help construct the inverse of $\mathbf{A}$ (if it exists), calculate the eigenvalues of $\mathbf{A}$, and determine the volume of the parallelepiped spanned by the columns of $\mathbf{A}$ (via its absolute value).
It can be determined recursively as

$$\det(\mathbf{A}) = \sum_{k=1}^{n} (-1)^{k+\ell} [\mathbf{A}]_{k,\ell} M_{k,\ell} = \sum_{k=1}^{n} (-1)^{k+\ell} [\mathbf{A}]_{\ell,k} M_{\ell,k} ,$$

where $\ell$ is a fixed integer in $1 \leq \ell \leq n$ and $M_{k,\ell}$ is the *minor* of $[\mathbf{A}]_{k,\ell}$, which is the determinant of the $(n-1) \times (n-1)$ submatrix formed by deleting the $k$-th row and $m$-th column of $\mathbf{A}$. (The determinant of a scalar is the scalar itself.)

**Properties:**

- If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ and $\mathbf{C} \triangleq \mathbf{AB}$, then $\det(\mathbf{C}) = \det(\mathbf{A}) \det(\mathbf{B})$.
- If $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are $m \times m$, $m \times n$, $n \times m$, and $n \times n$, respectively, then

$$\det\left(\begin{bmatrix} \mathbf{A} & \mathbf{0}_{m \times n} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}\right) = \det\left(\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0}_{n \times m} & \mathbf{D} \end{bmatrix}\right) = \det(\mathbf{A}) \det(\mathbf{D}) .$$

- If $\mathbf{A}$ is an $n \times n$ triangular matrix (either upper or lower), then

$$\det(\mathbf{A}) = \prod_{k=1}^{n} [\mathbf{A}]_{k,\ell} .$$

# Matrix Inverse

The inverse of an $n \times n$ square matrix $\mathbf{A}$, denoted $\mathbf{A}^{-1}$, is one for which

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n \, .$$

If such a matrix exists, $\mathbf{A}$ is said to be invertible. Otherwise, $\mathbf{A}$ is said to be singular. It can be shown that $\mathbf{A}$ is invertible if and only if $\det(\mathbf{A}) \neq 0$. In this case,

$$\left[\mathbf{A}^{-1}\right]_{k,\ell} = \frac{1}{\det(\mathbf{A})} (-1)^{\ell+k} M_{\ell,k} \, , \; 1 \leq k \leq n, 1 \leq \ell \leq n \, ,$$

where $M_{\ell,k}$ is the minor of $[\mathbf{A}]_{\ell,k}$. The quantity $C_{\ell,k} \triangleq (-1)^{\ell+k} M_{\ell,k}$ is the *cofactor* of $[\mathbf{A}]_{\ell,k}$.

## Matrix Inversion Lemma:

If $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are $m \times m$, $m \times n$, $n \times m$, and $n \times n$, respectively, and $\mathbf{A}$ and $\mathbf{D}$ are nonsingular, then

$$\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1} \, .$$

This lemma can be proved by considering the inverse of the block matrix $\mathbf{M}$ given by

$$\mathbf{M} \triangleq \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} .$$

In this case, the matrix $\mathbf{S}_{\mathbf{D};\mathbf{M}} \triangleq \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ is called the *Schur complement* of $\mathbf{D}$ in $\mathbf{M}$.

Similarly, $\mathbf{S}_{\mathbf{A};\mathbf{M}} \triangleq \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ is the Schur complement of $\mathbf{A}$ in $\mathbf{M}$.

# Eigenvalues & Eigenvectors

For an $n \times n$ square matrix $\mathbf{A}$, an $n \times 1$ nonzero vector $\mathbf{v}$ such that

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$$

is said to be an *eigenvector* of $\mathbf{A}$ with *eigenvalue* $\lambda$.
The eigenvalues of $\mathbf{A}$ can be obtained as the roots of its *characteristic polynomial* $p(\lambda)$ given by

$$p(\lambda) \triangleq \det(\lambda \mathbf{I}_n - \mathbf{A}) \ .$$

Note that there are exactly $n$ eigenvalues (counting multiplicity).
**Properties of eigenvalues & eigenvectors:**

- If $\{\lambda_1, \ldots, \lambda_n\}$ denote the set of eigenvalues of $\mathbf{A}$, then it can be shown that

$$\det(\mathbf{A}) = \prod_{k=1}^{n} \lambda_k \ , \ \operatorname{tr}(\mathbf{A}) = \sum_{k=1}^{n} \lambda_k \ .$$

- The eigenvalues of a triangular matrix (either upper or lower) are the diagonal elements.
- Suppose $\mathbf{A}$ is $n \times n$ with $n$ linearly independent eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$. If $\mathbf{V} \triangleq \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$ and $\mathbf{\Lambda} \triangleq \operatorname{diag}\left(\begin{bmatrix} \lambda_1 & \cdots & \lambda_n \end{bmatrix}^T\right)$, then we have

$$\mathbf{A} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^{-1} \ ,$$

  and we say that $\mathbf{A}$ is *diagonalizable*.
- If the eigenvalues of a matrix $\mathbf{A}$ are distinct, then $\mathbf{A}$ is diagonalizable. Otherwise, $\mathbf{A}$ may or may not be diagonalizable.

# Useful Miscellaneous Identities

- **Conjugate, transpose, and conjugate transpose:**
Suppose that $\alpha$ and $\beta$ are scalars, that $\mathbf{U}$ is $k \times \ell$, $\mathbf{V}$ is $k \times \ell$, $\mathbf{X}$ is $m \times n$, and $\mathbf{Y}$ is $n \times p$, and that $\mathbf{A}$ is $m_1 \times n_1$, $\mathbf{B}$ is $m_1 \times n_2$, $\mathbf{C}$ is $m_2 \times n_1$, and $\mathbf{D}$ is $m_2 \times n_2$. Then we have

$$(\alpha\mathbf{U} + \beta\mathbf{V})^* = \alpha^*\mathbf{U}^* + \beta^*\mathbf{V}^* \;,\; (\mathbf{XY})^* = \mathbf{X}^*\mathbf{Y}^* \;,\; \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^* = \begin{bmatrix} \mathbf{A}^* & \mathbf{B}^* \\ \mathbf{C}^* & \mathbf{D}^* \end{bmatrix} .$$

$$(\alpha\mathbf{U} + \beta\mathbf{V})^T = \alpha\mathbf{U}^T + \beta\mathbf{V}^T \;,\; (\mathbf{XY})^T = \mathbf{Y}^T\mathbf{X}^T \;,\; \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^T = \begin{bmatrix} \mathbf{A}^T & \mathbf{C}^T \\ \mathbf{B}^T & \mathbf{D}^T \end{bmatrix} .$$

$$(\alpha\mathbf{U} + \beta\mathbf{V})^\dagger = \alpha^*\mathbf{U}^\dagger + \beta^*\mathbf{V}^\dagger \;,\; (\mathbf{XY})^\dagger = \mathbf{Y}^\dagger\mathbf{X}^\dagger \;,\; \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^\dagger = \begin{bmatrix} \mathbf{A}^\dagger & \mathbf{C}^\dagger \\ \mathbf{B}^\dagger & \mathbf{D}^\dagger \end{bmatrix} .$$

- **Trace:**
Suppose that $\alpha$ and $\beta$ are scalars, that $\mathbf{U}$ is $k \times k$, $\mathbf{V}$ is $k \times k$, $\mathbf{X}$ is $m \times n$, and $\mathbf{Y}$ is $n \times m$, and that $\mathbf{A}$ is $m \times m$, $\mathbf{B}$ is $m \times n$, $\mathbf{C}$ is $n \times m$, and $\mathbf{D}$ is $n \times n$. Then we have

$$\mathrm{tr}(\alpha\mathbf{U} + \beta\mathbf{V}) = \alpha\mathrm{tr}(\mathbf{U}) + \beta\mathrm{tr}(\mathbf{V}) \;,\; \mathrm{tr}(\mathbf{XY}) = \mathrm{tr}(\mathbf{YX}) \;,\; \mathrm{tr}\left(\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}\right) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{D}) \;.$$

- **Inverse:**
Suppose that $c$ is a nonzero scalar and that $\mathbf{A}$ and $\mathbf{B}$ are invertible $n \times n$ matrices. Then we have

$$(c\mathbf{A})^{-1} = \frac{1}{c}\mathbf{A}^{-1} \;,\; \left(\mathbf{A}^{-1}\right)^{-1} = \mathbf{A} \;,\; (\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} .$$

$$\left(\mathbf{A}^T\right)^{-1} = \left(\mathbf{A}^{-1}\right)^T \;,\; \left(\mathbf{A}^\dagger\right)^{-1} = \left(\mathbf{A}^{-1}\right)^\dagger \;,\; \det\left(\mathbf{A}^{-1}\right) = \frac{1}{\det(\mathbf{A})} .$$

# Unitary and Normal Matrices

- **Unitary Matrices:**
  A matrix $\mathbf{U} \in \mathbb{C}^{m \times n}$ (with $m \geq n$) is said to be *unitary* if $\mathbf{U}^{\dagger}\mathbf{U} = \mathbf{I}_n$. Similarly, if $\mathbf{U} \in \mathbb{R}^{m \times n}$, then $\mathbf{U}$ is unitary if $\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$. If $m = n$, then we have
  $$\mathbf{U}^{\dagger}\mathbf{U} = \mathbf{U}\mathbf{U}^{\dagger} = \mathbf{I}_m \,, \; \left(\mathbf{U} \in \mathbb{C}^{m \times m}\right).$$
  $$\mathbf{U}^{T}\mathbf{U} = \mathbf{U}\mathbf{U}^{T} = \mathbf{I}_m \,, \; \left(\mathbf{U} \in \mathbb{R}^{m \times m}\right).$$

- **Normal Matrices:**
  A matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ is said to be *normal* if $\mathbf{A}^{\dagger}\mathbf{A} = \mathbf{A}\mathbf{A}^{\dagger}$. Similarly, if $\mathbf{A} \in \mathbb{R}^{n \times n}$, then $\mathbf{A}$ is normal if $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T$. It can be shown that $\mathbf{A}$ is normal if and only if it is diagonalizable by a unitary matrix. More specifically, $\mathbf{A}$ is diagonalizable if and only if
  $$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\dagger} \,, \; \left(\mathbf{A} \in \mathbb{C}^{n \times n}\right).$$
  $$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{T} \,, \; \left(\mathbf{A} \in \mathbb{R}^{n \times n}\right).$$
  Here, $\mathbf{U}$ is unitary and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix of eigenvalues of $\mathbf{A}$.

# Symmetric and Hermitian Matrices

- **Symmetric matrices:**
  An $n \times n$ matrix $\mathbf{A}$ is said to be *symmetric* if $\mathbf{A} = \mathbf{A}^T$.
  The set of $n \times n$ *real* symmetric matrices is a special one and will be denoted $\mathbb{S}^n$.
  Any real symmetric matrix $\mathbf{A}$ is *normal* and as such, is unitarily diagonalizable, i.e., admits a decomposition of the form $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ as described above.
- **Hermitian matrices:**
  An $n \times n$ matrix $\mathbf{A}$ is said to be *Hermitian* if $\mathbf{A} = \mathbf{A}^\dagger$.
  The set of $n \times n$ Hermitian matrices is a special one and will be denoted $\mathbb{H}^n$.
  As any real symmetric matrix is also Hermitian (yet there are Hermitian matrices which are not real symmetric), it follows that $\mathbb{S}^n \subset \mathbb{H}^n$.
  Any Hermitian matrix is *normal* and as such, is unitarily diagonalizable, i.e., admits a decomposition of the form $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$ as described above.
- **Properties of real symmetric and Hermitian matrices:**
  - The eigenvalues of any real symmetric or any Hermitian matrix are always real.
  - If $\mathbf{A} \in \mathbb{S}^n$, $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{B} \in \mathbb{H}^n$, and $\mathbf{w} \in \mathbb{C}^n$, then we have

$$\mathbf{v}^T\mathbf{A}\mathbf{v} \in \mathbb{R}\,,\ \mathbf{w}^\dagger\mathbf{B}\mathbf{w} \in \mathbb{R}\,.$$

# Positive Semidefinite / Positive Definite Matrices

- **Positive semidefinite matrices:**
  An $n \times n$ real symmetric matrix $\mathbf{A}$ is said to be *positive semidefinite* if
  $$\mathbf{v}^T \mathbf{A} \mathbf{v} \geq 0 \text{ for all } \mathbf{v} \in \mathbb{R}^n \,.$$
  Similarly, an $n \times n$ Hermitian matrix $\mathbf{B}$ is said to be positive semidefinite if
  $$\mathbf{w}^\dagger \mathbf{B} \mathbf{w} \geq 0 \text{ for all } \mathbf{w} \in \mathbb{C}^n \,.$$
  In either case, we will write $\mathbf{A} \succeq \mathbf{0}$ and $\mathbf{B} \succeq \mathbf{0}$.
  The sets of positive semidefinite real symmetric and Hermitian matrices are special ones and will be denoted $\mathbb{S}_+^n$ and $\mathbb{H}_+^n$, respectively.
- **Positive definite matrices:**
  An $n \times n$ real symmetric matrix $\mathbf{A}$ is said to be *positive definite* if
  $$\mathbf{v}^T \mathbf{A} \mathbf{v} > 0 \text{ for all } \mathbf{v} \in \mathbb{R}^n, \mathbf{v} \neq \mathbf{0} \,.$$
  Similarly, an $n \times n$ Hermitian matrix $\mathbf{B}$ is said to be positive definite if
  $$\mathbf{w}^\dagger \mathbf{B} \mathbf{w} > 0 \text{ for all } \mathbf{w} \in \mathbb{C}^n, \mathbf{w} \neq \mathbf{0} \,.$$
  In either case, we will write $\mathbf{A} \succ \mathbf{0}$ and $\mathbf{B} \succ \mathbf{0}$.
  The sets of positive semidefinite real symmetric and Hermitian matrices are special ones and will be denoted $\mathbb{S}_{++}^n$ and $\mathbb{H}_{++}^n$, respectively.

# Properties of Positive Semidefinite/Definite Matrices

- (*Definiteness of Diagonal Elements and Eigenvalues:*) The diagonal entries of a positive semidefinite (definite) matrix are always nonnegative (positive). A real symmetric or Hermitian matrix is positive semidefinite (definite) if and only if all of the eigenvalues are nonnegative (positive).
- (*Partial Ordering:*) For arbitrary square matrices $\mathbf{A}$ and $\mathbf{B}$, we will write $\mathbf{A} \succeq \mathbf{B}$ if $(\mathbf{A} - \mathbf{B}) \succeq \mathbf{0}$ and write $\mathbf{A} \succ \mathbf{B}$ if $(\mathbf{A} - \mathbf{B}) \succ \mathbf{0}$. This defines a *partial ordering* on the set of all square matrices.
- (*Matrix Square Roots:*) If $\mathbf{A} \in \mathbb{S}_{+}^{n}$, $\rho_{\mathbf{A}} = \mathrm{rank}(\mathbf{A})$, $\mathbf{B} \in \mathbb{H}_{+}^{n}$, and $\rho_{\mathbf{B}} = \mathrm{rank}(\mathbf{B})$, then there exists a $\mathbf{P} \in \mathbb{R}^{\rho_{\mathbf{A}} \times n}$ and $\mathbf{Q} \in \mathbb{C}^{\rho_{\mathbf{B}} \times n}$ such that
$$\mathbf{A} = \mathbf{P}^{T}\mathbf{P}\,,\ \mathbf{B} = \mathbf{Q}^{\dagger}\mathbf{Q}\,.$$
- (*Cholesky Decomposition:*) If $\mathbf{A} \in \mathbb{S}_{++}^{n}$ and $\mathbf{B} \in \mathbb{H}_{++}^{n}$, then there exist lower triangular matrices $\mathbf{L}_{\mathbf{A}} \in \mathbb{R}^{n \times n}$ and $\mathbf{L}_{\mathbf{B}} \in \mathbb{C}^{n \times n}$ with strictly positive diagonal entries such that
$$\mathbf{A} = \mathbf{L}_{\mathbf{A}}\mathbf{L}_{\mathbf{A}}^{T}\,,\ \mathbf{B} = \mathbf{L}_{\mathbf{B}}\mathbf{L}_{\mathbf{B}}^{\dagger}\,.$$

# Inner Products

One way to measure the *correlation* or *coherence* between two vectors or matrices is through the use of an *inner product*. An inner product $\langle \mathbf{x}, \mathbf{y} \rangle$ maps two vectors or matrices $\mathbf{x}$ and $\mathbf{y}$ (defined over a field $\mathbb{F}$) to the underlying field $\mathbb{F}$ and satisfies the following properties.

- *Conjugate symmetry:*

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle^* \ .$$

- *Linearity in the first argument:*

$$\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha \langle \mathbf{x}, \mathbf{z} \rangle + \beta \langle \mathbf{y}, \mathbf{z} \rangle \text{ for all } \alpha, \beta \in \mathbb{F} \, .$$

- *Positive definiteness:*

$$\langle \mathbf{x}, \mathbf{x} \rangle \geq 0 \text{ with equality if and only if } \mathbf{x} = \mathbf{0} \, .$$

**Common inner products:**

$$\langle \mathbf{x}, \mathbf{y} \rangle = \begin{cases} \mathbf{y}^T\mathbf{x} \, , & \text{for } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \text{ (standard inner product on } \mathbb{R}^n\text{)} \\ \mathbf{y}^\dagger\mathbf{x} \, , & \text{for } \mathbf{x}, \mathbf{y} \in \mathbb{C}^n \text{ (standard inner product on } \mathbb{C}^n\text{)} \\ \mathbf{y}^\dagger\mathbf{P}\mathbf{x} \, , & \text{for } \mathbf{x}, \mathbf{y} \in \mathbb{C}^n, \, \mathbf{P} \in \mathbb{H}^n_{++} \end{cases} \, .$$

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \begin{cases} \mathrm{tr}\left(\mathbf{Y}^T\mathbf{X}\right) \, , & \text{for } \mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m\times n} \text{ (standard inner product on } \mathbb{R}^{m\times n}\text{)} \\ \mathrm{tr}\left(\mathbf{Y}^\dagger\mathbf{X}\right) \, , & \text{for } \mathbf{X}, \mathbf{Y} \in \mathbb{C}^{m\times n} \text{ (standard inner product on } \mathbb{C}^{m\times n}\text{)} \\ \mathrm{tr}\left(\mathbf{Y}^\dagger\mathbf{P}\mathbf{X}\right) \, , & \text{for } \mathbf{X}, \mathbf{Y} \in \mathbb{C}^{m\times n}, \, \mathbf{P} \in \mathbb{H}^n_{++} \end{cases} \, .$$

# Norms: Definition and Examples of Vector Norms

One way to measure the length of a vector or matrix in some sense is through the use of a *norm*. A norm $||\mathbf{x}||$ maps a vector or matrix $\mathbf{x}$ (defined over a field $\mathbb{F}$) to $\mathbb{R}_+$ and satisfies the following.

- *Positive definiteness:*

$$||\mathbf{x}|| \geq 0 \text{ with equality if and only if } \mathbf{x} = \mathbf{0} \,.$$

- *Homogeneity:*

$$||\alpha \mathbf{x}|| = |\alpha| \cdot ||\mathbf{x}|| \ \text{ for all } \alpha \in \mathbb{F} \,.$$

- *Triangle inequality:*

$$||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}|| \ \text{ for all } \mathbf{x}, \mathbf{y} \,.$$

**Common vector norms:**
For all examples considered here, we assume $\mathbf{x} \in \mathbb{F}^n$ and $x_k = [\mathbf{x}]_k$.

- *Euclidean norm: ($\ell_2$-norm)*

$$||\mathbf{x}||_2 = \sqrt{\sum_{k=1}^{n} |x_k|^2} = \sqrt{\mathbf{x}^\dagger \mathbf{x}} \,.$$

- *Chebyshev norm: ($\ell_\infty$-norm)*

$$||\mathbf{x}||_\infty = \max \left\{ |x_1| , \ldots , |x_n| \right\} \,.$$

- *$\ell_p$-norm:*

$$||\mathbf{x}||_p = \left( \sum_{k=1}^{n} |x_k|^p \right)^{\frac{1}{p}} \quad \text{(valid only for } p \geq 1) \,.$$

# Matrix Norms: Operator and Entrywise Norms

There are conventionally three varieties of matrix norms: operator norms, entrywise norms, and Schatten norms (which will be discussed at another time).

- **Operator Norms:**
  If $||\cdot||_{\rm a}$ and $||\cdot||_{\rm b}$ are vector norms defined over $\mathbf{F}^m$ and $\mathbf{F}^n$, respectively, then the *operator norm* of $\mathbf{X} \in \mathbb{F}^{m \times n}$, induced by $||\cdot||_{\rm a}$ and $||\cdot||_{\rm b}$, is defined as

  $$||\mathbf{X}||_{\rm a,b} \triangleq \sup\left\{||\mathbf{X}\mathbf{u}||_{\rm a} : ||\mathbf{u}||_{\rm b} \leq 1\right\} .$$

  *Example:* When both $||\cdot||_{\rm a}$ and $||\cdot||_{\rm b}$ are the Euclidean norm, the resulting operator norm is called the *spectral norm* or the $\ell_2$-norm and is given by

  $$||\mathbf{X}||_2 = \sqrt{\lambda_{\max}\left(\mathbf{X}^\dagger\mathbf{X}\right)} ,$$

  where $\lambda_{\max}\left(\mathbf{X}^\dagger\mathbf{X}\right)$ denotes the maximum eigenvalue of $\mathbf{X}^\dagger\mathbf{X}$.
- **Entrywise Norms:**
  If we treat the matrix $\mathbf{X} \in \mathbb{F}^{m \times n}$ as a vector of size $mn$ and apply a familiar vector norm, we obtain an *entrywise norm*. Assuming $X_{k,\ell} = [\mathbf{X}]_{k,\ell}$, some examples are as follows.

  $$||\mathbf{X}||_F = \sqrt{\sum_{k=1}^{m}\sum_{\ell=1}^{n}\left|X_{k,\ell}\right|^2} = \sqrt{\mathrm{tr}\left(\mathbf{X}^\dagger\mathbf{X}\right)} \quad \text{(Frobenius norm)} .$$

  $$||\mathbf{X}||_p = \left(\sum_{k=1}^{m}\sum_{\ell=1}^{n}\left|X_{k,\ell}\right|^p\right)^{\frac{1}{p}} \quad (\ell_p\text{-norm}) .$$

# The Dual Norm

The concept of duality occurs frequently throughout the study of convex optimization. One way in which duality manifests itself is through the *dual norm*. If $||\mathbf{z}||$ is some norm for either a vector or matrix $\mathbf{z}$, the associated dual norm $||\mathbf{z}||_{\circledast}$ is defined as follows.

$$||\mathbf{z}||_{\circledast} \triangleq \sup\{\text{Re}[\langle \mathbf{x}, \mathbf{z} \rangle] : ||\mathbf{x}|| \leq 1\} \ .$$

The dual norm can be expressed in the following equivalent forms, which are more convenient for analysis.

$$||\mathbf{z}||_{\circledast} = \sup\{|\langle \mathbf{x}, \mathbf{z} \rangle| : ||\mathbf{x}|| = 1\} = \sup\left\{\left|\left\langle \frac{\mathbf{x}}{||\mathbf{x}||}, \mathbf{z} \right\rangle\right| : \mathbf{x} \neq 0\right\} \ .$$

**Properties:**

- The dual norm $||\mathbf{z}||_{\circledast}$ is indeed a norm.
- The primal norm and dual norm satisfy the inequality
$$||\mathbf{x}|| \cdot ||\mathbf{y}||_{\circledast} \geq |\langle \mathbf{x}, \mathbf{y} \rangle| = |\langle \mathbf{y}, \mathbf{x} \rangle| \text{ for all } \mathbf{x}, \mathbf{y} \ .$$
- The dual of the dual norm, denoted $||\mathbf{z}||_{\circledast\circledast}$, is the original norm $||\mathbf{z}||$, i.e.,
$$||\mathbf{z}||_{\circledast\circledast} = ||\mathbf{z}|| \ .$$

# Examples of Dual Norms

To calculate the dual norm $||\mathbf{z}||_{\circledast}$, typically an upper bound on $|\langle \mathbf{x}, \mathbf{z} \rangle|$ is computed and $\mathbf{x}$ is chosen so as to achieve the upper bound, if possible. Using this approach leads to the following examples of dual norms.

| **Primal Norm:** | **Dual Norm:** |
|---|---|
| *Euclidean norm* – $\sqrt{\sum_{k=1}^{n} \|z_k\|^2}$ | *Euclidean norm* – $\sqrt{\sum_{k=1}^{n} \|z_k\|^2}$ |
| $\ell_1$*-norm* – $\sum_{k=1}^{n} \|z_k\|$ | $\ell_\infty$-norm – $\max\{\|z_1\|, \ldots, \|z_n\|\}$ |
| $\ell_p$*-norm* – $\left( \sum_{k=1}^{n} \|z_k\|^p \right)^{\frac{1}{p}}$ | $\ell_q$*-norm* – $\left( \sum_{k=1}^{n} \|z_k\|^q \right)^{\frac{1}{q}}$ $(q = p/(p-1))$ |
| *spectral norm* – $\|\|\mathbf{Z}\|\|_2 = \sqrt{\lambda_{\max}(\mathbf{Z}^\dagger \mathbf{Z})}$ | *nuclear norm* – $\|\|\mathbf{Z}\|\|_{2\circledast} = \mathrm{tr}\left(\sqrt{\mathbf{Z}^\dagger \mathbf{Z}}\right)$ |