MATHEMATICAL FOUNDATIONS OF SIGNAL PROCESSING Fall 2014
Benjamín Béjar Haro, Mihailo Kolundžija, Reza Parhizkar, & Martin Vetterli
Teaching assistant: Hanjie Pan

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Review of some mathematical tools

1 Vector spaces, Matrices

- A *vector space* defined over a field $\mathbb{F}$ (most often $\mathbb{R}$ or $\mathbb{C}$) is a set V together with two binary operations (multiplication and addition), satisfying
  1. **Associativity**. $v_1 + (v_2 + v_3) = (v_1 + v_2) + v_3$,
  2. **Commutativity**. $v_1 + v_2 = v_2 + v_1$,
  3. **Existence of additive identity**. $\exists \mathbf{0} \in V$, such that $v + \mathbf{0} = v$,
  4. **Existence of additive inverse**. $\forall v \in V,\ \exists (-v) \in V$ such that $v + (-v) = \mathbf{0}$,
  5. **Distributivity with respect to vector addition**. $a(v_1 + v_2) = av_1 + av_2,\ a \in \mathbb{F}$,
  6. **Distributivity with respect to scalar addition**. $(a_1 + a_2)v = a_1 v + a_2 v,\ a_1, a_2 \in \mathbb{F}$,
  7. **"Associativity" of scalar multiplication**. $a_1(a_2 v) = (a_1 a_2)v,\ a_1, a_2 \in \mathbb{F}$,
  8. **Existence of (scalar–vector) multiplicative identity**. $\exists 1 \in \mathbb{F}$ such that $1v = v,\ \forall v \in V$.

  Ex: $\mathbb{R}^n$, $\mathbb{C}^n$, $\mathbb{R}^\infty$, $L^2([0,1])$ are all vector spaces over $\mathbb{R}$ or over $\mathbb{C}$ with the usual scalar–vector multiplication and vector (function in the case of $L^2([0,1])$) addition.

- A *subspace* W of V, is a subset of V which is itself a vector space.

  Remark: If $W \subset V$ and V is a vector space, then it is sufficient to check whether $aw \in W$ and $w_1 + w_2 \in W$ for W to be a subspace, where $a \in \mathbb{F}$, $w_1, w_2 \in W$.

- The *span* $\text{span}\,(S)$ of a set of vectors S is the smallest vector space containing S.

  Remark: In a strict sense, a vector space is closed under finite addition. Therefore the span of an infinite set of vectors S is an open subspace. It is usually more convenient to deal with its closure $\overline{\text{span}}\,(\cdot)$.

  Ex: Consider the infinite Cartesian basis $E = \{\delta_n\}_{n\in\mathbb{Z}}$ and the all-1 sequence $z = (\ldots, 1, 1, 1, 1, 1, \ldots)$. Then $z \notin \text{span}\,(E)$, but $z \in \overline{\text{span}}\,(E)$.

- A vector space is *countable* if it is spanned by a countable set of vectors

  Ex: $\ell^2(\mathbb{Z})$ is countable, $L^2([0,\ 1])$ is countable, but $L^2(\mathbb{R})$ is not countable.

- The *dimension* of a vector space, $\dim(V)$, is the minimal number of vectors spanning V.

  Ex: $\dim(\mathbb{R}^n) = n$, $\dim(\mathbb{C}^\infty) = \infty$, $\dim\left(\text{span}\left([1,\ 2,\ 3]^T, [2,\ 4,\ 6]^T\right)\right) = 1$.

- A *matrix* A can be functionally defined as a *linear operator between countable vector spaces* (defined over the same field $\mathbb{F}$):

$$A : V_1 \to V_2\ , \quad A(\alpha x + \beta y) = \alpha Ax + \beta Ay,$$

  where $\alpha,\ \beta \in \mathbb{F}$ and $x_1,\ x_2 \in V_1$.

As a convention vectors are assumed to be column vectors.

2 Hermitian Transpose

The *Hermitian transpose* of a matrix A, is noted A^*. The matrix entries a_{ij} are such that

$$a^*_{ij} = a_{ji},$$

where "$*$" denotes the conjugate symmetric ($A_{i,j}$ is a scalar). The *reverse order law* states:

$$(AB)^* = B^* A^*.$$

3 Range, Nullspace

Let A be a matrix of dimension $m \times n$ defined on the scalar field $\mathbb{F}$, i.e. $A \in \mathbb{F}^{m \times n}$:

- The *range*, $\mathcal{R}(A)$ is the set of all vectors y that can be "produced" by a matrix A, $\mathcal{R}(A) = \{y \ : \ y = Ax, x \in \mathbb{F}^n\}$. Range is a *linear subspace* of $\mathbb{F}^m$.
- The *nullspace* or the *kernel* $\mathcal{N}(A)$ is the set of solutions x to $Ax = 0$. It is also a linear subspace.
  If $\mathcal{N}(A) = \{0\}$, it is called a trivial nullspace.

4 Rank, Dimension

- The dimensions of a matrix A are its number of rows m and number of columns n.

  Remark: Do not confuse the column dimension of a matrix with the dimension of its range! Make sure you understand the difference between matrix dimensions and a subspace dimension (such as range and nullspace dimensions).
- The *rank* of a matrix A is the dimension of its range, i.e. the cardinality of the largest set of linearly independent columns of A.
- For a matrix A with column dimension n:

$$\dim(\mathcal{N}(A)) + \dim(\mathcal{R}(A)) = n.$$

- The rank of a matrix A is equal to the rank of its Hermitian transpose A^*, i.e., the number independent columns and independent rows are the same for any matrix A.
- As a corollary, $\dim(\mathcal{R}(A)) \leq \min(m, n)$.
- An $m \times n$ matrix A is *full-rank* if $\dim(\mathcal{R}(A)) = \min(m, n)$.

5 Inner-product, Adjoint

An *inner product* on a vector space V over C is a mapping $\langle \cdot, \cdot \rangle : V^2 \to \mathbb{C}$ satisfying

$$\begin{aligned} \langle \cdot, \cdot \rangle \ : \ & V \times V \to \mathbb{C}. \\ & \langle \alpha(u+v), w \rangle = \alpha \langle u, w \rangle + \alpha \langle v, w \rangle, \\ & \langle u, u \rangle \geq 0 \ , \quad \langle u, u \rangle = 0 \Leftrightarrow u = 0, \\ & \langle u, v \rangle = \langle v, u \rangle^*. \end{aligned}$$

The *usual inner product* in $\mathbb{C}^n$ (or $\mathbb{R}^n$):

$$\langle u, v \rangle = v^* u.$$

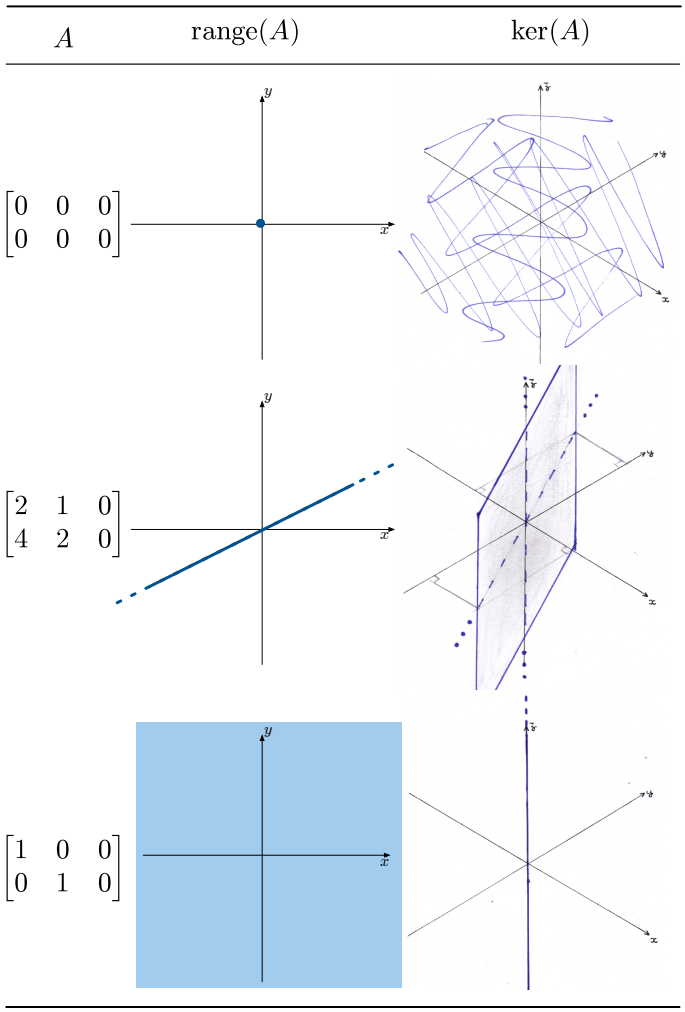

Figure 1: Range and kernel for some matrices.

The usual inner product is always implied if not otherwise mentioned. To differentiate between different inner products, the vector space is sometimes included in the notation $\langle \cdot, \cdot \rangle_V$. A vector space with an inner-product is called an inner-product space.

An inner-product induces a norm. For example the usual inner product induces the Euclidean norm:

$$\|u\| = \sqrt{\langle u, u \rangle}$$

The *adjoint* of a matrix $A : V_1 \to V_2$ is given by its Hermitian transpose $A^* : V_2 \to V_1$. The term adjoint stresses its operational properties. Let $A : V_1 \to V_2$, $u \in V_2$, $v \in V_1$:

$$\langle u, Av \rangle_{V_2} = \langle A^* u, v \rangle_{V_1}.$$

6 Singularity, Inverse

- A matrix is *singular* iff its nullspace is non-trivial.
- A matrix A has a proper *inverse* iff $A : V \to V$ is a bijection from/to the same vector space (A^{-1} is the inverse of A, also A is the inverse of A^{-1}):

$$A^{-1}A = AA^{-1} = I.$$

7 Pseudo-inverse

- Any matrix A has a *pseudo-inverse* $A^\dagger$. It is defined by the following four properties:

$$\begin{array}{rcl|rcl} AA^\dagger A & = & A & (AA^\dagger)^* & = & AA^\dagger \\ A^\dagger AA^\dagger & = & A & (A^\dagger A)^* & = & A^\dagger A \end{array}$$

- There are special cases where the pseudo-inverse can be seen as a proper inverse lacking commutativity, i.e. as a lateralized inverse:
  - Iff A has linearly independent columns (A is nonsingular) A^*A is invertible. Therefore the pseudo-inverse is a *left-inverse* $A^\dagger A = I$, with $A^\dagger = (A^*A)^{-1}A^*$.

    Remark: Do not confuse pseudo-inverses with left-inverses and their explicit formulation. Left-inverses are only a particular case!
  - Iff A has linearly independent rows (A^* is nonsingular) AA^* is invertible. Therefore the pseudo-inverse is a *right-inverse* $AA^\dagger = I$, with $A^\dagger = A^*(AA^*)^{-1}$.

8 Square Matrices

Let A be an $n \times n$ matrix over a field $\mathbb{F}$.

8.1 Determinant

Determinants are often abusively used to study singularity. We will not give computational formula (found in any textbook), but rather give its intuition. From this intuition it will be clear that determinants should be used only if a notion of "volume" is involved.

The *determinant* $\det(A)$ of a square matrix A can be interpreted as the *oriented volume* of the hyper-parallelepiped defined by the column vectors of A. The "orientation" is defined by the number of coordinate inversions compared to the canonical basis (each inversion flips the sign).

Ex: Determinants of 2×2 *matrix.*

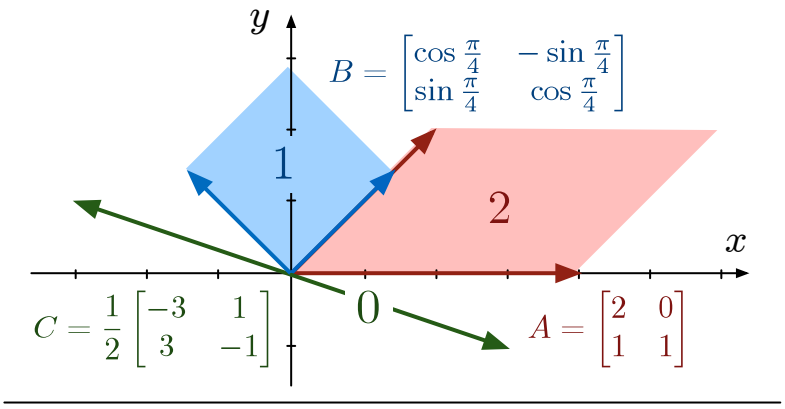

Figure 2: Matrix determinants

This intuitive geometric definition suggests that rotations and reflections R have a unit determinant. Indeed, $|\det(R)| = 1$ for rotation and reflection matrices.

Ex: The use of the Jacobian (determinant) in change of measure becomes very clear. It ensures that the volume of the infinitesimal area element is the same before and after the transformation. For example going from cartesian to polar coordinates:

$$\iint_C dx\,dy = \iint_P \left| \begin{matrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \phi} \end{matrix} \right| dr\,d\phi$$

8.2 Characteristic equation/polynomial

The *characteristic equation* of A is:

$$Ax = \lambda x\ , \quad x \in V_{\overline{\mathbb{F}}},\ x^*x = 1,\ \lambda \in \overline{\mathbb{F}},$$

where $\overline{\mathbb{F}}$ is the algebraic closure of $\mathbb{F}$.

The characteristic equation has for solution (λ, x) if λ verifies:

$$\det(\lambda I - A) = 0.$$

The left-hand side of the equation is called the *characteristic polynomial* and has degree n. The *fundamental theorem of algebra* states that a polynomial of degree n has exactly n roots in an algebraically closed field (counting multiple roots).

Remark: This does not hold on non algebraically closed fields, e.g. a polynomial of degree 2 in $\mathbb{R}$ may have 0, 1 or 2 roots.

Ex: Let $A \in \mathbb{R}^{2\times 2}$, the characteristic equation is

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} x = \lambda x.$$

There is no real solution. However as expected from the fundamental theorem of algebra, there are 2 complex solutions ($\mathbb{C}$ is the algebraic closure of $\mathbb{R}$):

$$(\lambda_0, x_0) = (1 + j, [1\ \ -i]^T),\ (\lambda_1, x_1) = (1 - j, [1\ i]^T).$$

A pair (λ, x) solution of the characteristic equation is called an eigenpair of A, with λ the eigenvalue and x the eigenvector. The set of all eigenvalues is called the spectrum of A.

8.3 Diagonalization, Schur decomposition

A square matrix A is *diagonalizable* if $\exists S \in \mathbb{F}^{n\times n}$ and a diagonal matrix $\Lambda \in \mathbb{F}^{n\times n}$ such that:

$$A = S\Lambda S^{-1}$$

This decomposition is called the *eigenvalue decomposition* of A. The link with the eigenvalues and eigenvectors, and thus with the characteristic equation, is as follows:

$$A = S\Lambda S^{-1} \Leftrightarrow AS = S\Lambda \Leftrightarrow As_j = \lambda_j s_j,\ \forall j = 1\ldots,\ n.$$

Ex: Not all matrices are diagonalizable, for example $R = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ *is not diagonalizable.*

For all square matrix $A \in \mathbb{F}^{n\times n}$, there exist $U \in \mathbb{F}^{n\times n}$ unitary ($U^{-1} = U^*$, see next section) and $R \in \mathbb{F}^{n\times n}$ an upper triangular matrix such that:

$$A = URU^*.$$

This is called the *Schur decomposition* of A.

Ex: Our previous non diagonalizable example $R = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ *has a trivial Schur decomposition:* $R = IRI^*$.

8.4 Special square matrices

- *Unitary matrices* have their adjoint for inverse: $U^{-1} = U^*$. The eigenvalues of a unitary matrix have unit modulus: $|\lambda_j| = 1$. Unitary matrices are isometries (with respect to the norm induced by the inner-product):

  $$\|Ux\| = \|x\|.$$

- *Normal matrices* commute with their adjoint:

  $$AA^* = A^*A.$$

  Normal matrices are at the core of the *spectral theorem*:

  $$A \in \mathbb{F}^{n\times n} \text{ normal } \Rightarrow A = U\Lambda U^*,$$

  with U unitary and Λ diagonal. Therefore the Schur decomposition and eigenvalue decomposition coincide for normal matrices.
- *Hermitian/self-adjoint matrices* are their own adjoint: $H = H^*$. Real and complex Hermitian matrices have real eigenvalues, and Hermitian symmetry implies normality.
- *Positive semi-definite/definite matrices* have a non-negative/strictly positive real eigenvalues, $\lambda_j \in \mathbb{R}^+$ for positive semi-definite matrices and $\lambda_j \in \mathbb{R}^+/\{0\}$. Concepts of matrix definiteness are typically defined for Hermitian matrices, and we will only use them for Hermitian matrices (even if there are options for non-Hermitian). Positive semi-definite (definite) matrices are usually defined by the relation

  $$x^*Ax \begin{matrix} \geq \\ > \end{matrix} 0\ , \quad \forall x \in V\{0\},$$

  which makes explicit positive definite matrices are suitable to define an inner-product

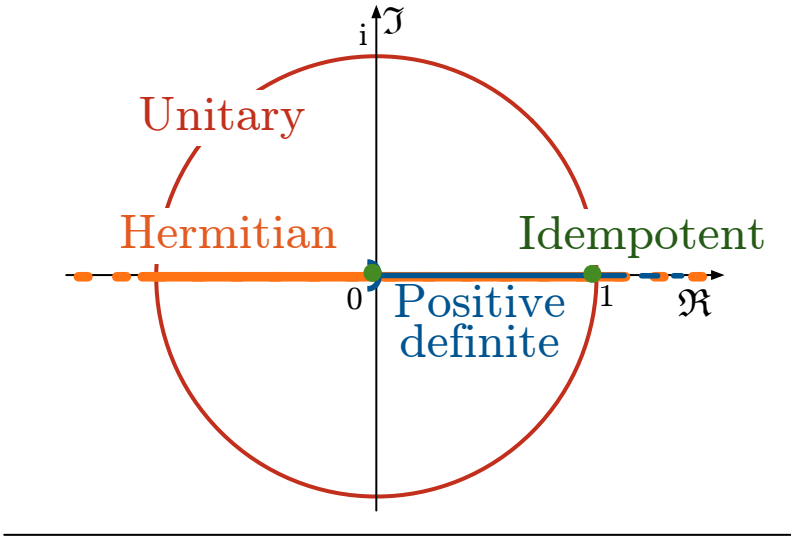

Figure 3: Characterization of square matrices by their spectrum (eigenvalues).

- *Idempotent matrices* verify $P = P^2$, i.e. the transformation P applied n times to x is the same as applying it a single time. They are also called *projections*. The eigenvalues of an idempotent matrix can only be 0 or 1.
  Idempotent and self-adjoint matrices are *orthogonal projections*. *Remark:* Idempotent have real eigenvalues but they may not be Hermitian, e.g. $P = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$. An idempotent matrix is Hermitian iff it is normal.
- *Circulant matrices* implement a circular convolution, e.g.:

$$C = \begin{bmatrix} a & b & c & d \\ d & a & b & c \\ c & d & a & b \\ b & c & d & a \end{bmatrix}.$$

  They are diagonalized by the DFT matrix W (which is a unitary transform): $C = W\Lambda W^*$. The spectrum of C is the DFT of the convolution filter $[d\ c\ b\ a]$.
- *Toeplitz and Hankel matrices* have very similar properties. A Toeplitz matrix has constant coefficients along the diagonals and implements a linear convolution, e.g.:

$$T = \begin{bmatrix} a & b & c & d \\ e & a & b & c \\ f & e & a & b \\ g & f & e & a \end{bmatrix}, \ t_{i,j} = t_{i-j}.$$

  Hankel matrices are constant along the antidiagonals, and can thus be obtained by multiplying a Toeplitz matrix by the exchange matrix J (mirror of the identity matrix). Hankel matrices implement a time-reversed linear convolution. Their properties are therefore very similar to those of Toeplitz matrices, and results on Toeplitz matrices almost always extend to Hankel matrices. The sequence $[g\ f\ e\ a\ b\ c\ d]$ created from the first column/line of T is called the *generator* of T.

  Ex: Toeplitz matrices are connected to stationarity (see Yule-Walker equations for more).

Ex: A very practical theorem by Grenander and Szegö states that under some conditions (a generator with "fast enough" decay) large Toeplitz matrices converge to a circulant matrix in term of their spectrum and energy. They are thus asymptotically diagonalized by the DFT matrix W.

9 Rectangular matrices

Let $A \in \mathbb{F}^{m \times n}$,

- If $m > n$, the matrix is called *tall*. If its columns are linearly independent, it has a left inverse.
- If $n > m$, the matrix is called *fat*. If its rows are linearly independent, it has a right inverse.

9.1 The singular value decomposition

The *Singular Value Decomposition* (SVD) holds for any square and rectangular matrix. It has the form:

$$A = USV^*,$$

where $U \in \mathbb{F}^{m \times m}$ and $V \in \mathbb{F}^{n \times n}$ are unitary matrices, and $S \in \mathbb{F}^{m \times n}$ is a block matrix with a Hermitian diagonal block $S_1 = \mathrm{diag}(\sigma_1, \ \ldots, \ \sigma_{\min(m,n)})$ and a 0 block.

By convention, the sign of the left and right singular vectors are chosen such that For example, take $m > n$ (A is tall):

$$\left[\begin{array}{c|c} & \\ U_1 & U_0 \\ & \end{array}\right] \left[\begin{array}{c} \begin{matrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{matrix} \\ \hline 0_{(m-n)\times n} \end{array}\right] \left[\begin{array}{c} \\ V \\ \\ \end{array}\right]$$

This to the so called *thin SVD*:

$$A = U_1 S_1 V^*.$$

The vectors U_1 and U_0 respectively form an orthonormal basis for $\mathcal{R}(A)$ and $\mathcal{N}(A)$. The same holds for fat matrices by Hermitian transposition.

The SVD is linked to the eigenvalue decomposition in the sense that U and V are respectively the eigenvectors of the normal matrices AA^* and A^*A. As for the eigenvalue decomposition, *singular triplets* $(u_j, \ v_j, \ \sigma_j)$ are solution to a characteristic equation:

$$\left[\begin{array}{c} A \\ \hline A^* \end{array}\right] \left[\begin{array}{c} v_j \\ \hline u_j \end{array}\right] = \sigma_j \left[\begin{array}{c} u_j \\ \hline v_j \end{array}\right].$$

The SVD is useful to compute the pseudo-inverse:

$$A^\dagger = VS^\dagger U^*,$$

$S^\dagger$ is S^T with inverted singular values (σ_i replaced by $\frac{1}{sigma_i}$).

10 Best approximation (Least squares)

The system $Ax = y$ has an exact solution iff $y \in \mathcal{R}(A)$. If so, and if A has

- independent columns: $x = A^\dagger y$ is the unique solution.
- non independent columns, there are infinitely many solutions of the form $x = A^\dagger y + a^\perp$, where $a^\perp$ belongs to the nullspace of A. The solution with $a^\perp = 0$ has the least energy.

  Ex: The previous case (independent columns) fits this definition, however the nullspace is in this case trivial, which makes the solution unique.
- independent lines and columns (A is square and invertible): $x = A^{-1}y$.

However many practical systems do not have an exact solution. In this case an approximate solution optimizing some criterion is chosen.

The best approximation in the LS sense $\widehat{x}$ minimizes $\|A\widehat{x} - y\|_2$. It is obtained with the pseudo-inverse of A:

$$\widehat{x} = A^\dagger y.$$

See the SVD section for an always valid formulation of the pseudo-inverse.

Ex: Solving in the LS sense provides the maximum likelihood (ML) estimate in case A is noiseless and y is corrupted by some additive white gaussian noise (AWGN).

11 Building projections

The pseudo-inverse can be used to build orthogonal projections:

- $AA^\dagger$ is the orthogonal projection into $\mathcal{R}(A)$.
- $A^\dagger A$ is the orthogonal projection into $\mathcal{R}(A^*)$, the row-space of A.
- For U an orthonormal set of vectors, $U^\dagger = U^*$, therefore projection into the range of an orthonormal set of vectors U is given by the outer-product UU^*.
- Orthogonal projection into the nullspace of A is $(I - A^\dagger A)$, i.e. the complement of the orthogonal projection into $\mathcal{R}(A^*)$

12 The Gram-Schmidt process

Given a matrix A, the *Gram-Schmidt process* iteratively builds an orthonormal set of vectors S spanning $\mathcal{R}(A)$. Let $A = [a_1 \ \cdots \ a_n]$, from the index $j = 1$ and a set $S = \emptyset$:

- Orthogonalize a_j against all vectors in S:

$$a_j^\perp = a_j - \sum_{s \in S} \langle a_j, s \rangle s.$$

- Normalize $a_j^\perp$: $a_j^\perp \leftarrow a_j^\perp / \|a_j^\perp\|$.
- Add this new vector to the set S: $S \leftarrow S \cup \{a_j^\perp\}$.
- repeat for $j \leftarrow j + 1$ until $j = n$.

By construction S is an orthonormal set of vectors, and since any of the column vectors of A can be obtained as linear combination of vectors of S and vice-versa, they span the same subspace.