H6: unitary similarity and unitary equivalence

6.1 unitary and orthogonal groups	
def: unitary group $\mathcal{U}(n)$	= the set of matrixes U in $\mathbf{C}^{nxn}$ satisfying: $$U^H U = I_n = U U^H$$ > subgroup of GL(n,ℂ) and |det(U)| = 1
def: special unitary group $S\mathcal{U}(n)$	= unitary group for which every U∈$S\mathcal{U}(n)$ det(U) = +1 > subgroup of SL(n,ℂ)
$\mathcal{U}(n)$ a compact set	finite-dimensional unitary matrices form a compact set
def: orthogonal group O(n)	= the restriction of the unitary group to real numbers
def: special orth. group SO(n)	= orthogonal group of matrices with det(O) = +1
6.2 elementary unitary transformations	
6.2.1 permutation matrices	
def: permutation matrix	For a permutation $\sigma \in S_n$ > the associated permutation matrix $P(\sigma) \in \mathbb{R}^{nxn}$ is an orthogonal matrix with nonzero entries given by: $P^i{}_j(\sigma) = \delta^i_{\sigma(j)}$.
6.2.2 Givens transformations	
def: Givens transformation	= unitary matrix $G \in \mathbb{F}^{nxn}$ specified by two parameters c,s∈𝔽 satisfying $\|c\|^2+\|s\|^2 = 1$ and two integers $1 \leq i \leq j \leq n$ such that the nonzero entries of G are given by: $$G^k{}_k = 1 \text{ for } k \neq i, k \neq j, \qquad G^i{}_i = G^j{}_j = c, \qquad G^j{}_i = \bar{s}, \qquad G^i{}_j = -s.$$
6.2.3 Householder transformation	
def: Householder transformation	In an inner product space V, this corresponds to a unitary self-adjoint operator which maps for **u** a given unit vector: $$\boldsymbol{w} \to \boldsymbol{w} - 2\langle \boldsymbol{u}, \boldsymbol{w}\rangle \boldsymbol{u},$$ In standard space $\mathbb{F}^n$ this corresponds to H(**v**): $$H = I - \frac{2}{\boldsymbol{v}^H \boldsymbol{v}} \boldsymbol{v}\boldsymbol{v}^H$$
6.2.4 Discrete Fourier transformation	
def: discrete Fourier transform	for a function $f : \{0, \ldots, n-1\} \to \mathbb{C} : j \to f_j$ is given by $$F_k = \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} f_j \exp\left(-\mathrm{i}\frac{2\pi}{n} jk\right), \qquad \forall k = 0, \ldots, n-1.$$
prop: inverse of discr. Ft	the inverse of the discrete Fourier transform is given by: $$f_j = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} F_k \exp\left(+\mathrm{i}\frac{2\pi}{n} jk\right)$$
def: circulant	= a matrix $A \in \mathbb{C}^{nxn}$ with entries: $A^i{}_j = f_{(j-i) \bmod n}$.
prop: circulant and DFT	*A circulant matrix* $A \in \mathbb{F}^{n\times n}$ *is diagonalised by the Fourier transform matrix* $U \in \mathbb{F}^{n\times n}$, *i.e.* $AU = U\Lambda$ *with* $\Lambda = \mathrm{diag}(\lambda_0, \lambda_1, \ldots, \lambda_{n-1})$, *where the eigenvalues are given by* $$\lambda_k = \sum_{j=0}^{n-1} f_j \omega^{kj} = \sum_{j=0}^{n-1} f_j e^{-\mathrm{i}\frac{2\pi}{n}kj}. \qquad (6.11)$$

6.3 QR decomposition revisited	
QR-decomp. via Householder reflections	A first Householder reflection H_1 is applied based on the first column ie: use $w = (A^i{}_1)_{i=1,\dots,m}$ in order to cancel the elements in rows i=2,...,m in 1st column > now construct a second Hh refection H_2 acting on the second column ie: use $w = ((H_1^H A)^i{}_2)_{i=2,\dots,m}$, to zero out elements in rows i=3,...,m in the in 2nd column > this transformation doesn't effect the first column >> do this for every column > find a sequence of Hh reflections that eliminate the entries at rows i=k+1,...m: $H_1, H_2, \dots, H_n$, where H_k acts on rows $k, \dots, m$ We can then obtain: $H_{n-1}^H \dots H_2^H H_1^H A = \tilde{R} \implies A = \underbrace{H_1 H_2 \dots H_{n-1}}_{\tilde{Q}} \tilde{R}.$
6.4 Schur decomposition and power iteration	
6.4.1 Schur decomposition	
th: Schur triangulation	Any matrix $A \in \mathbb{F}^{n\times n}$ admits a Schur decomposition which takes the form: $A = ZTZ^H$ *where* $Z \in U(n)$ *and* $T \in \mathbb{F}^{n\times n}$ *is upper triangular.*[6] > this isn't unique for a certain matrix A ie: you can fiend multiple Schur decompositions
6.4.2 normal matrices revisited	
prop: schur decomp. on a normal matrix	For A a normal matrix > its Schur decompositions, the upper triangular matrix T is diagonal ie: A is diagonalised by a unitary matrix and the Schur and eigenvalue decomp. coincide
6.4.3 power method and subspace iteration	
def: power iteration	For a square matrix A > start from a given vector $v_0 = v$ and iterate over following steps for k=0,1,2,... 1. $q_k = v_k / \lVert v_k \rVert$ (normalisation); 2. $v_{k+1} = A q_k$; 3. $\mu_k = \langle q_k, v_{k+1} \rangle = q_k^H A q_k$.
prop: convergence of pow. it.	For $A \in \mathbb{F}^{n\times n}$ For λ_1 the eigenvalue for A for which: - $\lvert\lambda_1\rvert = \rho_A$ ie: λ_1 is the largest eigenvalue - λ is simple or semisimple (1) For P_1 a projector onto the eigenspace $V_{\lambda 1}$ along the direct sum of all eigenspaces and generalised eigenspaces For any vector V for which $P_1 v \neq 0$ > the power method will converge with: $\mu_\infty = \lim_{k\to\infty} \mu_k = \lambda_1$ *and* $q_\infty = \lim_{k\to\infty} q_k$ *a normalised eigenvector in* V_{λ_1}. (1) its geometric and algebraic multiplicity coincide and there are no non-trivial Jordan blocks associated with λ_1

def: subspace iteration	For a matrix $A \in \mathbb{F}^{n \times n}$ > this starts from m≤n vectors that are collected as the columns of a matrix $V_0 \in \mathbb{F}^{n \times m}$ and iterates over the following steps over k=0,1,2,... 1. $Q_k R_k = V_k$ (QR decomposition); 2. $V_{k+1} = AQ_k$; 3. $T_k = Q_k^H V_{k+1} = Q_k^H A Q_k$. If the subspace iteration method converges, so that each of the sequences Q_k, V_k, R_k and T_k reach a fixed point which we denote as Q_∞, V_∞, R_∞ and T_∞ respectively, we find $$V_\infty = AQ_\infty = Q_\infty R_\infty \quad \text{and} \quad T_\infty = R_\infty. \tag{6.17}$$
def: QR algorithm	For a matrix $A \in \mathbb{F}^{n \times n}$ > this starts from T_0 = A and Z_0=I and iterates: 1. $Q_k R_k = T_{k-1}$; 2. $Z_k = Z_{k-1} Q_k$; 3. $T_k = R_k Q_k$. It can easily be seen that for any iteration, we have $$A = T_0 = Q_1 R_1 = Q_1 T_1 Q_1^H = \ldots = (Q_1 Q_2 \cdots Q_k) T_k (Q_1 Q_2 \cdots Q_k)^H = Z_k T_k Z_k^H$$
6.5 bilinear and quadratic forms revisited	
matrix congruence ~C	like linear matrices with similarity transforms aka $\tilde{A} = TAT^{-1}$, a bilinear matrix C can be transformed to ~C via: $$\tilde{C} = T^{-H} C T^{-1}$$
unitary similarity transform Ã	$\tilde{A} = UAU^H$ lies in the intersection of similarity and congruence > because $U^{-1} = U^H$
6.5.1 signature and Sylvester's law	
prop: general canonical form	The matrix representation B of a symmetric bilinear form B is congruent with a canonical form given by: $$V^T B V = \mathrm{diag}(\underbrace{+1,\ldots,+1}_{n_+ \text{ times}}, \underbrace{-1,\ldots,-1}_{n_- \text{ times}}, \underbrace{0,\ldots,0}_{n_0 \text{ times}}). \tag{6.19}$$ *The three numbers (n_+, n_-, n_0) are sometimes referred to as the* ***inertia*** *or* ***signature*** *of the symmetric matrix* B.
properties of n	n_0 = nullity of B = ν(B) $n_+ + n_-$ = rank of B = ρ(B) > thus $n_0 + n_+ + n_- = n = \dim(B)$
positive (semi)definite-ness via inertia	Bilinear form is: - positive semidefinite if $n_- = 0$ - positive definite if $n_+ = n$
principle axis	= the basisvectors $\mathbf{u}_k$ ~ $\mathbf{v}_k$ which diagonalise the quadratic form > eigenvalues d_k have meaning dependent on the chosen basis
th: Sylvester's law 6.7	Two symmetric matrices B, $\sim B \in \mathbb{R}^{n \times n}$ have the same number of n_+, n_- and n_0 if and only if they are congruent ie: there is some $T \in GL(n, \mathbb{R})$ such that $B = T^T \tilde{B} T$.

6.5.2 Cholesky decomposition and Lagrange reduction	
Cholesky decomposition	For a real symmetric matrix B > the LDU decomp. reduces B = LDL^T > in LDU all the diagonal elements of L are 1 Modify this, such that D only contains the sign of the diagonal elements and absolute value of nonzero diagonal elements is absorbed > by rescaling the columns of L and with a square root of this absolute value For the associated quadratic form: $$q(\boldsymbol{x}) = \boldsymbol{x}^{\mathsf{T}}\mathsf{B}\boldsymbol{x} = \sum_{i,j=1}^{n} B_{ij}x^i x^j = B_{11}(x^1)^2 + 2B_{12}x^1x^2 + \dots,$$ this procedure is equivalent to that of *Lagrange reduction*, whereby one sequentially completes the squares. One combines the term $B_{11}(x^1)^2$ with all the terms $2B_{1k}x^1x^k$ for $k = 2, \dots, n$ into a square $$\operatorname{sgn}(B_{11})(\sqrt{\lvert B_{11}\rvert}x^1 + \sum_{k=2}^{n} B_{1k}/\sqrt{\lvert B_{11}\rvert}x^k)^2,$$ Thereby subtracting the additional terms that were needed for this square > variable x^1 has now been completely eliminated > repeat this process for x^2,x^3,... > until the quadratic form is rewritten as pure squares of ne variables: $$y^j = \textstyle\sum_{k=j}^{n}(\mathsf{L}^{\mathsf{T}})^j{}_k x^k.$$ >> if B is positive definite, ie: $n_+ = n$, its canonical form corresponds to I_n > thus B = LL^T
6.6 singular value and polar decomposition	
6.6.1 unitary equivalence	
unitarily equivalent	= matrices related via $\tilde{A} = U^H AV$ with unitary U and V > is equivalence relation
6.6.2 singular value decomposition	
prop: singular value decomp.	For a matrix $A \in \mathbb{F}^{m\times n}$ > there exist unitary matrices U∈U(m) and V∈U(n) such that: $$\mathsf{A} = \mathsf{USV}^{\mathsf{H}}$$ with $S \in \mathbb{F}^{m\times n}$ has only nonzero elements on the diagonal > can be made positive and ordered such that: $$S^1{}_1 = \sigma_1 \geq S^2{}_2 = \sigma_2 \geq \ldots \geq S^p{}_p = \sigma_p > 0 \; for\ some\ p \leq \min(m,n)$$ these $(\sigma_1,\dots,\sigma_p)$ are known as the *singular values* of A
properties of singular values	- singular values of a matrix A are unique - matrices U and V which decompose A aren't
singular value decomp. of a real matrix A	The singular value decomp. of a real matrix A decomposes it into the following steps: • A rotation to a new set of orthonormal basis vectors (V) • A rescaling of the basis vectors / coordinates (S) • Another rotation, possibly with reflection, to the final basis (U).
equations with singular values	*Remark* 6.38. If we denote the ith column of U and V as $\boldsymbol{u}_i \in \mathbb{F}^m$ and $\boldsymbol{v}_i \in \mathbb{F}^n$ respectively for $i = 1, \dots, p$, then we can write $$\mathsf{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i, \qquad \mathsf{A}^{\mathsf{H}}\boldsymbol{u}_i = \sigma_i \boldsymbol{v}_i. \qquad (6.22)$$

singular value decomp. of square matrices	For A∈$\mathbb{F}^{nxn}$ a square matrix of size n > we can relate the singular value decomp. of A to the eigenvalue decomp. of a Hermitian matrix of size 2nx2n: $\begin{bmatrix} O & A \\ A^H & O \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} U & U \\ V & -V \end{bmatrix} \right) = \left(\frac{1}{\sqrt{2}} \begin{bmatrix} U & U \\ V & -V \end{bmatrix} \right) \begin{bmatrix} S & O \\ O & -S \end{bmatrix}$ >> can be used for singular value decomp
singular value decomp for non-square matrices	first compute a QR-decomp of A and then use the previous for the singular value decomp > we obtain: $A = QR = QU_R S_R V_R^H = U_A S_A V_A^H,$ and thus: $U_A = QU_R,\ S_A = S_R \text{ and } V_A = V_R.$
def: thin singular value decomp	if m≠n, a different yet equivalent decomp of A is given by: $A = U_k S_k V_k^H \text{ with } k = \min(m, n)$
def: compact singular value decomposition	if only p<min(m,n) singular values are non-zero, then: $A = U_p S_p V_p^H$
def: rank r truncated singular value decomposition	if we reduce k to some value r<p and discard the p-r smallest non-zero singular values > no longer an equality, but consider as an approximation of A: $U_r S_r V_r^H$,
6.6.3 rank, norm, condition number	
prop: link between decomp and rank, norm, image	1: The singular value decomposition exposes the rank of the matrix: > ρ(A) = p, the number of singular values 2: U_p, the column of U corresponding to the non-zero singular values: > constitute an orthonormal basis for im(A) 3: the columns k=p+1,…,n of V provide an orthonormal basis for ker(A)
def: full rank	= matrix A∈$\mathbb{F}^{mxn}$ for which the rank has its maximal value p=ρ(A)=min(m,n) > all other matrices are *rank deficient*
prop: norm and σ_1	For A∈$\mathbb{F}^{mxn}$ a linear map between $\mathbb{F}^n$ and $\mathbb{F}^m$ with Euclidean norm $\Vert . \Vert_2$ > now it holds: $\Vert A \Vert_2 = \sigma_1,$
prop: Frobenius norm and σ	*The Frobenius norm of* $A \in \mathbb{F}^{m \times n}$ *is given by* $\Vert A \Vert_F = \sqrt{\sum_{i=1}^p \sigma_i^2}.$
prop: condition number and σ	*The condition number of an invertible matrix* $A \in \mathbb{F}^{n \times n}$ *is given by* $\kappa(A) = \sigma_1 / \sigma_n.$
6.6.4 least squares and pseudo-inverses	
def: minimum norm least squares solution	For a general matrix A∈$\mathbb{F}^{mxn}$ For L: $L = \{x \in \mathbb{F}^n \mid \Vert Ax - y \Vert_2 = \min_{x' \in \mathbb{F}^n} \Vert Ax' - y \Vert_2\}$ denote the set of all solutions that minimise the norm of the residual > then: $x^* = \arg\min_{x \in L} \Vert x \Vert$ is known as the minimum norm least squares solution
prop: solution of min. norm least squares solution	For a general matrix A∈$\mathbb{F}^{mxn}$ > the min. norm least squares solution for A**x** = **y** is uniquely given by: $x^* = V_p S_p^{-1} U_p^H y,$
def: Moore-Penrose pseudoinverse	For a general matrix A∈$\mathbb{F}^{mxn}$ this is: $A^+ = V_p S_p^{-1} U_p^H \quad (6.25)$ in terms of the compact singular value decomposition $A = U_p S_p V_p^H$, where $p = \rho(A)$.

6.6.5 low rank approximations	
th: Eckart-Young-Mirsky theorem	For $A \in \mathbb{F}^{mxn}$ > the rank r matrix B that minimises $\|A-B\|_2$ for 2-norm is given by: $B = U_r S_r V_r^H$
th: Eckart-Young-Mirsky theorem for Frobenius norm	For $A \in \mathbb{F}^{mxn}$ > the rank r matrix B that minimises $\|A-B\|_2$ for F-norm is given by: $B = U_r S_r V_r^H$
6.6.6 polar decomposition	
prop: polar decomposition	For $A \in \mathbb{F}^{mxn}$ with m≥n > it can be decomposed as: $$A = UP \qquad (6.26)$$ *where* $U \in \mathbb{F}^{m \times n}$ *is isometric* ($U^H U = I_n$) *and* $P \in \mathbb{F}^{n \times n}$ *is Hermitian and positive semidefinite.* >> generalised: $z = e^{i \arg(z)} \|z\|$
prop: isometric factor	*Given* $A \in \mathbb{F}^{m \times n}$ *with* $m \geq n$, *the isometric matrix* B *that minimises* $\|A - B\|_F$ *is given by the isometric factor in the polar decomposition of* A.