Chapter 2

Here we learned about norms on spaces of matrices. Two cases are here important:

i) Induced matrix norms

Definition 0.0.1. *Let* $\mathbf{A} \in \mathbb{K}^{m\times n}$ *and we equip* $\mathbb{K}^n$ *with* $\|\cdot\|_{(n)}$ *and* $\mathbb{K}^m$ *with* $\|\cdot\|_{(m)}$. *The induced matrix norm is then*

$$\|\mathbf{A}\|_{(m,n)} = \sup_{\substack{\mathbf{x}\in\mathbb{K}^n \\ \mathbf{x}\neq 0}} \frac{\|\mathbf{A}\mathbf{x}\|_{(m)}}{\|\mathbf{x}\|_{(n)}} = \sup_{\substack{\mathbf{x}\in\mathbb{K}^n \\ \|\mathbf{x}\|_{(n)}=1}} \|\mathbf{A}\mathbf{x}\|_{(m)} \tag{1}$$

ii) Matrix norm on the vector space of matrices

Definition 0.0.2. *A function* $\|\cdot\| : \mathbb{K}^m \to \mathbb{R}$ *is called a norm if*

1. $\|\mathbf{x}\| \geqslant 0$ *for all* $\mathbf{x} \in \mathbb{K}^m$ *and* $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$
2. $\|\mathbf{x}+\mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|$ *for all* $\mathbf{x},\mathbf{y} \in \mathbb{K}^m$
3. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ *for all* $\mathbf{x} \in \mathbb{K}^m$ *and for all* $\alpha \in \mathbb{K}$.

We have also seen central statements like

- Young's product inequality:

Lemma 0.0.1 (Young's product inequality)**.** *Let* $a, b \in \mathbb{R}_{\geqslant 0}$. *Then*

$$ab \leqslant \frac{1}{p}a^p + \frac{1}{q}b^q \tag{2}$$

for $1 \leqslant p, q \leqslant \infty$ *and* $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. Let $a, b \in \mathbb{R}_{\geqslant 0}$, and $t = \frac{1}{p}$ and $1 - t = \frac{1}{q}$. Then

$$\ln(ta^p + (1-t)b^q) \underset{(*)}{\geqslant} t\ln(a^p) + (1-t)\ln(b^q) = \ln(a) + \ln(b) = \ln(ab) \tag{3}$$

where we used that ln in concave in $(*)$. □

- Hölder inequality:

Theorem 0.1 (Hölder inequality)**.** *Let* $\mathbf{x},\mathbf{y} \in \mathbb{K}^m$. *Then*

$$|\langle\mathbf{x},\mathbf{y}\rangle| \leqslant \|\mathbf{x}\|_p\|\mathbf{y}\|_q, \tag{4}$$

where $1 \leqslant p, q \leqslant \infty$ *and* $\frac{1}{p} + \frac{1}{q} = 1$.

- Cauchy–Schwarz inequality:

Corollary 0.1.1 (Cauchy–Schwarz inequality)**.** *For* $p, q = 2$ *the Hölder inequality yields*

$$|\langle\mathbf{x},\mathbf{y}\rangle| \leqslant \|\mathbf{x}\|_2\|\mathbf{y}\|_2. \tag{5}$$

- Minkowski inequality

  Theorem 0.2 (Minkowski inequality)**.** *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{K}^m$ *and* $p \geqslant 1$*. Then*

$$\|\mathbf{x} + \mathbf{y}\|_p \leqslant \|\mathbf{x}\|_p + \|\mathbf{y}\|_p \tag{6}$$

- The standard inner product on matrix spaces:

  Definition 0.2.1. *Let* $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{m\times n}$*. The standard inner product of matrices is defined as*

$$\langle \mathbf{A}, \mathbf{B} \rangle = \operatorname{Tr}(\mathbf{A}^*\mathbf{B}), \tag{7}$$

- The Frobenius norm is induced by the standard inner product:

  Proposition 0.2.1. *The Frobenius norm is induced by the standard inner product of matrices, i.e., for* $\mathbf{A} \in \mathbb{K}^{m\times n}$

$$\|\mathbf{A}\|_{\mathrm{F}} = \sqrt{\langle \mathbf{A}, \mathbf{B} \rangle}. \tag{8}$$

- Unitarily invariant norms:

  Theorem 0.3. *For any* $\mathbf{A} \in \mathbb{K}^{m\times n}$ *and unitary* $\mathbf{U} \in \mathbb{K}^{m\times m}$*, we have*

$$\|\mathbf{U}\mathbf{A}\|_2 = \|\mathbf{A}\|_2 \quad \text{and} \quad \|\mathbf{U}\mathbf{A}\|_{\mathrm{F}} = \|\mathbf{A}\|_{\mathrm{F}} \tag{9}$$

  Proof. Let $\mathbf{A} \in \mathbb{K}^{m\times n}$ and $\mathbf{U} \in \mathbb{K}^{m\times m}$ be unitary. Then

$$\|\mathbf{U}\mathbf{A}\mathbf{x}\|_2 = \sqrt{\mathbf{x}^*\mathbf{A}^*\mathbf{U}^*\mathbf{U}\mathbf{A}\mathbf{x}} = \|\mathbf{A}\mathbf{x}\|_2 \Rightarrow \|\mathbf{U}\mathbf{A}\|_2 = \|\mathbf{A}\|_2 \tag{10}$$

  and

$$\|\mathbf{U}\mathbf{A}\|_{\mathrm{F}} = \sqrt{\operatorname{Tr}((\mathbf{U}\mathbf{A})^*\mathbf{U}\mathbf{A})} = \sqrt{\operatorname{Tr}(\mathbf{A}^*\mathbf{A})} = \|\mathbf{A}\|_{\mathrm{F}} \tag{11}$$

  □

In the homework assignments you have seen central statements like:

- Hermitian matrices have real-valued eigenvalues.
- Skew hermitian matrices have purely imaginary eigenvalues
- And matrix inequalities: $\|x\|_2 \leqslant \sqrt{m}\|x\|_\infty$

  Proof. Starting from the definition of the 2-norm, we find

$$\|x\|_2 = \left(\sum_{i=1}^m |x_i|^2\right)^{1/2} \leqslant \left(\sum_{i=1}^m \max_{i\in[m]} |x_i|^2\right)^{1/2} = \sqrt{m}\left(\max_{i\in[m]} |x_i|^2\right)^{1/2} = \sqrt{m}\left(\max_{i\in[m]} |x_i|\right),$$

  hence $\sqrt{m}\|x\|_\infty$. The inequality is sharp for $x = (1, 1, \ldots, 1)^\top$, i.e., the vector with all entries equal to one, since $\|x\|_2 = \sqrt{m}$ and $\|x\|_\infty = 1$. □

Chapter 3

The most central object of this course and large parts of numerical linear algebra; the singular value decomposition:

Definition 0.3.1. *Let $\mathbf{A} \in \mathbb{K}^{m \times n}$. We call the factorization*

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*, \tag{12}$$

where $\mathbf{U} \in \mathbb{K}^{m \times m}$ and $\mathbf{V} \in \mathbb{K}^{n \times n}$ are unitary, and $\mathbf{\Sigma} \in \mathbb{K}^{m \times n}$ is diagonal, singular value decomposition of $\mathbf{A}$.

Theorem 0.4. *Every matrix $\mathbf{A} \in \mathbb{K}^{m \times n}$ has a singular value decomposition and the singular values $\{\sigma_i\}$ are uniquely determined. Moreover, if $\mathbf{A}$ is square and σ_i distinct, the left and right singular vectors $\{\mathbf{u}_j\}$ and $\{\mathbf{v}_j\}$ are uniquely determined up to complex signs, i.e., complex scaling factors of length one.*

The proof is long but parts can be asked:

- Then A^*A is positive semi-definite, indeed,

$$\mathbf{x}^*\mathbf{A}^*\mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})^*(\mathbf{A}\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|_2 \geqslant 0. \tag{13}$$

Proposition 0.4.1. *Let $\mathbf{A} \in \mathbb{K}^{m \times n}$. Then $\|\mathbf{A}\|_2 = \sigma_{\max}(\mathbf{A})$, i.e., the largest singular value.*

Proof. Let $\mathbf{A} \in \mathbb{K}^{m \times n}$, with singular value decomposition $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ and σ_1 being the largest singular value. Then for $\|\mathbf{x}\|_2 = 1$

$$\|\mathbf{A}\mathbf{x}\|_2^2 = \langle \mathbf{x}, \mathbf{A}^*\mathbf{A}\mathbf{x} \rangle = \sum_{i=1}^{n} \sigma_i^2 \langle \mathbf{x}, \mathbf{v}_i\mathbf{v}_i^*\mathbf{x} \rangle \leqslant \sigma_1^2 \sum_{i=1}^{n} |\mathbf{v}_i^*\mathbf{x}|^2 = \sigma_1^2 \|\mathbf{V}^*\mathbf{x}\|^2 \leqslant \sigma_1^2 \|\mathbf{V}^*\|^2 = \sigma_1^2 \tag{14}$$

which is tight for $\mathbf{x} = \mathbf{v}_1$. □

We learned two key applications:

- Low-rank approximation
- Moore-Penrose inverse:

  Definition 0.4.1. *Let $\mathbf{A} \in \mathbb{K}^{m \times n}$. The matrix $A^+ \in \mathbb{K}^{n \times m}$ is called the pseudo inverse (Moose-Penrose) inverse of $\mathbf{A}$ if*

$$\begin{array}{llll} \text{i)} & \mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A} & \text{iii)} & (\mathbf{A}\mathbf{A}^+)^* = \mathbf{A}\mathbf{A}^+ \\ \text{ii)} & \mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+ & \text{iv)} & (\mathbf{A}^+\mathbf{A})^* = \mathbf{A}^+\mathbf{A} \end{array}$$

That have different properties:

Low-rank

Theorem 0.5 (Eckast-Young-Mirsky – spectral norm)**.** *Let* $\mathbf{A} \in \mathbb{K}^{m\times m}$ *with* $\operatorname{rank}(\mathbf{A}) = r$. *For any* k *with* $1 \leqslant k < r$, *define*

$$A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^*. \tag{15}$$

Then

$$\|\mathbf{A} - \mathbf{A}_k\|_2 = \inf_{\substack{\mathbf{B}\in\mathbb{K}^{m\times m} \\ \operatorname{rank}(\mathbf{B})\leqslant k}} \|\mathbf{A} - \mathbf{B}\|_2 = \sigma_{k+1} \tag{16}$$

Proof. First note that

$$\|A - A_k\|_2 = \sigma_{k+1}. \tag{17}$$

It remains to show that $\mathbf{A}_k$ is the infimum. To that end, assume the exist $\mathbf{B}_k = \mathbf{X}\mathbf{Y}^*$ where $\mathbf{X}, \mathbf{Y}$ have k-columns and that

$$\|\mathbf{A} - \mathbf{B}_k\|_2 < \|\mathbf{A} - \mathbf{A}_k\|_2 = \sigma_{k+1}. \tag{18}$$

However, since

$$\operatorname{rank}(\mathbf{Y}) = k < k + 1 = \operatorname{rank}([\mathbf{v}_1|\ldots|\mathbf{v}_{k+1}]) \tag{19}$$

there exists a linear combination of right singular vectors of

$$\mathbf{w} = c_1\mathbf{v}_1 + \ldots + c_{k+1}\mathbf{v}_{k+1} \tag{20}$$

with

$$\mathbf{Y}^*\mathbf{w} = \mathbf{0}. \tag{21}$$

W.l.o.g. we assume $\mathbf{w}$ is normalized, otherwise we normalize $\mathbf{w}$. Then,

$$\|\mathbf{A} - \mathbf{B}_k\|_2^2 \geqslant \|(\mathbf{A} - \mathbf{B}_k)\,\mathbf{w}\|_2^2 = \|A\mathbf{w}\|_2^2 = c_1^2\sigma_1^2 + \ldots + c_{k+1}^2\sigma_{k+1}^2 \geqslant \sigma_{k+1}^2 \tag{22}$$

□

Theorem 0.6 (Courant-Fisher min-max – singular values)**.** *For* $\mathbf{A} \in \mathbb{K}^{m\times n}$, *we have*

$$\sigma_k = \max_{\substack{V\subset\mathbb{K}^n \\ \dim(V)=k}} \min_{\substack{\|\mathbf{v}\|=1 \\ \mathbf{v}\in V}} \|\mathbf{A}\mathbf{v}\|_2 \tag{23}$$

and

$$\sigma_{k+1} = \min_{\substack{V\subset\mathbb{K}^n \\ \dim(V)=n-k}} \max_{\substack{\|\mathbf{v}\|=1 \\ \mathbf{v}\in V}} \|\mathbf{A}\mathbf{v}\|_2. \tag{24}$$

Theorem 0.7 (Weyl's inequality)**.** *Let* $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{m\times n}$ *and denote its singular values by* $\sigma_i(\mathbf{A})$ *and* $\sigma_i(\mathbf{B})$, *respectively. We then have*

$$\sigma_{i+j-1}(\mathbf{A} + \mathbf{B}) \leqslant \sigma_i(\mathbf{A}) + \sigma_j(\mathbf{B}). \tag{25}$$

Theorem 0.8 (Eckert-Young-Mirsky for Frobenins norm)**.** *Let* $\mathbf{A} \in \mathbb{K}^{m\times m}$ *with* $\operatorname{rank}(\mathbf{A}) = r$. *For any* k *with* $1 \leqslant k < r$, *define*

$$\mathbf{A}_k = \sum_{j=1}^{k} \sigma_j \mathbf{u}_j \mathbf{v}_j^*. \tag{26}$$

Then

$$\|\mathbf{A} - \mathbf{A}_k\|_F = \inf_{\substack{\mathbf{B}\in\mathbb{K}^{m\times m} \\ \operatorname{rank}(\mathbf{B})\leqslant k}} \|\mathbf{A} - \mathbf{B}\|_F = \sqrt{\sigma_{k+1}^2 + \ldots + \sigma_r^2}. \tag{27}$$

Moore Penrose inverse

Proposition 0.8.1. *Let $\mathbf{A} \in \mathbb{K}^{m\times n}$ with $m \leqslant n$ and $\mathbf{A}^+$ its Moore-Penrose inverse. Then*

$$\operatorname{range}(\mathbf{A}^+) \perp \ker(\mathbf{A}). \tag{28}$$

Proof. Let $\mathbf{A} \in \mathbb{K}^{m\times n}$ with $\mathbf{A}^+$ its Moore-Penrose inverses. Recall that

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A} \quad \text{and} \quad (\mathbf{A}^+\mathbf{A})^* = \mathbf{A}^+\mathbf{A}. \tag{29}$$

Moreover let $\mathbf{y} \in \operatorname{range}(\mathbf{A}^+)$, i.e., $\mathbf{y} = \mathbf{A}^+\mathbf{b}$ for some $\mathbf{b} \in \mathbb{K}^m$, and $\mathbf{x} \in \ker(\mathbf{A})$. Then

$$\langle \mathbf{y}, \mathbf{x}\rangle = \langle \mathbf{A}^+\mathbf{b}, \mathbf{x}\rangle = \langle \mathbf{A}^+\mathbf{A}\mathbf{A}^+\mathbf{b}, \mathbf{x}\rangle = \langle \mathbf{A}^+\mathbf{b}, (\mathbf{A}^+\mathbf{A})^*\mathbf{x}\rangle = \langle \mathbf{A}^+\mathbf{b}, \mathbf{A}^+\mathbf{A}\mathbf{x}\rangle = 0 \tag{30}$$

□

Theorem 0.9. *Let $\mathbf{A} \in \mathbb{K}^{m\times n}$, the Moore-penrose inverse $\mathbf{A}^+$ is unique.*

Proposition 0.9.1. *Let $\mathbf{A} \in \mathbb{K}^{m\times n}$ with $m > n$ and $\operatorname{rank}(\mathbf{A}) = n$. Then*

$$\mathbf{A}^+ = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*. \tag{31}$$

Theorem 0.10. *If $\mathbf{A} \in \mathbb{K}^{m\times m}$ attains an inverse, then $\mathbf{A}^{-1} = \mathbf{A}^+$.*

Proof. Note that

$$\mathbf{I} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{A}\mathbf{A}^+\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}\mathbf{A}^+ \tag{32}$$

hence

$$\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{I} = \mathbf{A}^{-1}\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+ \tag{33}$$

□

Theorem 0.11. *Let $\mathbf{A} \in \mathbb{K}^{m\times n}$ with $\mathbf{A}^+ \in \mathbb{K}^{n\times m}$ its pseudo inverse, then*

$$\left(\mathbf{A}^+\right)^+ = \mathbf{A}. \tag{34}$$

Application of MP inverse:

The MP inverse solves the over-determined least squares problem, i.e., minimize

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2. \tag{35}$$

where $\mathbf{A} \in \mathbb{K}^{m\times n}$ and $m \geqslant n$ – we say "$\mathbf{A}$ is tall and skinny". We have more equations than variables and consequently zero solutions to the system. We therefore seek $\mathbf{x} \in \mathbb{K}^n$ that minimizes the above residual, i.e.,

$$\min_{\mathbf{x}\in\mathbb{K}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2. \tag{36}$$

To that end, we compute the gradient of with respect to $\mathbf{x}$:

$$\frac{\partial}{\partial \mathbf{x}}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 = 2\mathbf{A}^*(\mathbf{A}\mathbf{x} - \mathbf{b}). \tag{37}$$

Enforcing first-order optimality yields the normal equation

$$\mathbf{A}^*\mathbf{A}\mathbf{x} = \mathbf{A}^*\mathbf{b}. \tag{38}$$

Assuming $\mathbf{A}^*\mathbf{A}$ is invertible, which holds if $\mathbf{A}$ has full rank, we can solve the normal equation, i.e.,

$$\mathbf{x} = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{b} = \mathbf{A}^+\mathbf{b}. \tag{39}$$

Chapter 4

This Chapter was all about QR factorization. We learned

Definition 0.11.1. *Let* $\mathbf{P} \in \mathbb{K}^{m\times m}$*. We call* $\mathbf{P}$ *a projector if and only if*

$$\mathbf{P}^2 = \mathbf{P}, \tag{40}$$

i.e., $\mathbf{P}$ *is idempotent.*

Remark 0.11.1. *This definition includes both, orthogonal and non-orthogonal projectors. To avoid confusion, we call non-orthogonal projectors oblique projectors.*

Proposition 0.11.1. *If* $\mathbf{P} \in \mathbb{K}^{m\times m}$ *is a projector, then* $\mathbf{I} - \mathbf{P}$ *is also a projector.*

Proof. Note that

$$(\mathbf{I} - \mathbf{P})^2 = (\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P}) = \mathbf{I} - 2\mathbf{P} + \mathbf{P}^2 = \mathbf{I} - \mathbf{P} \tag{41}$$

which shows the claim. □

Definition 0.11.2. *Let* $\mathbf{P} \in \mathbb{K}^{m\times m}$ *be a projector. We call* $\mathbf{P}$ *an orthogonal projector if and only if*

$$\langle \mathbf{P}\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{P}\mathbf{y}\rangle \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{K}^m, \tag{42}$$

i.e., $\mathbf{P} \in \mathbb{H}_m(\mathbb{K})$*.*

Definition 0.11.3. *Let* $\mathbf{A} \in \mathbb{K}^{m\times n}$*. We call the factorization*

$$\mathbf{A} = \mathbf{Q}\mathbf{R} \tag{43}$$

where $\mathbf{Q} \in \mathbb{K}^{m\times m}$ *unitary, and* $\mathbf{R} \in \mathbb{K}^{m\times n}$ *is an upper triangular matrix, a QR-factorization of* $\mathbf{A}$*.*

Remark 0.11.2. *We shall now take a closer look at the QR-factorization:*
Consider a reduced QR-factorization of $\mathbf{A} \in \mathbb{K}^{m\times n}$ *with* $n \leqslant m$*, i.e.,*

$$[\mathbf{a}_1|\dots|\mathbf{a}_n] = [\mathbf{q}_1|\dots|\mathbf{q}_n] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1_n} \\ 0 & r_{22} & \cdots & r_{2_n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{nn} \end{bmatrix} \tag{44}$$

hence

$$\begin{aligned}
\mathbf{a}_1 &= r_{11}\mathbf{q}_1 & &\Leftrightarrow & \mathbf{q}_1 &= \frac{\mathbf{a}_1}{r_{11}} = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|} \\
\mathbf{a}_2 &= r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2 & &\Leftrightarrow & \mathbf{q}_2 &= \frac{\mathbf{a}_2 - r_{12}\mathbf{q}_1}{r_{22}} = \frac{\mathbf{a}_2 - \langle \mathbf{q}_1, \mathbf{a}_2\rangle \mathbf{q}_1}{r_{22}} = \frac{(\mathbf{I} - \mathbf{q}_1\mathbf{q}_1^*)\mathbf{a}_2}{\|(\mathbf{I} - \mathbf{q}_1\mathbf{q}_1^*)\mathbf{a}_2\|} \\
\mathbf{a}_3 &= r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2 + r_{33}\mathbf{q}_3 & &\Leftrightarrow & \mathbf{q}_3 &= \frac{\mathbf{a}_3 - r_{13}\mathbf{q}_1 - r_{23}\mathbf{q}_2}{r_{33}} = \frac{(\mathbf{I} - \mathbf{q}_1\mathbf{q}_1^* - \mathbf{q}_2\mathbf{q}_2^*)\mathbf{a}_3}{\|(\mathbf{I} - \mathbf{q}_1\mathbf{q}_1^* - \mathbf{q}_2\mathbf{q}_2^*)\mathbf{a}_3\|} \\
& & &\vdots & & \\
\mathbf{a}_i &= \sum_{j=1}^{i} r_{ji}\mathbf{q}_j & &\Leftrightarrow & \mathbf{q}_i &= \frac{\mathbf{a}_i - \sum_{j=1}^{i-1} r_{ji}\mathbf{q}_j}{r_{ii}} = \frac{(\mathbf{I} - \sum_{j=1}^{i-1} \mathbf{q}_j\mathbf{q}_j^*)\mathbf{a}_i}{\|(\mathbf{I} - \sum_{j=1}^{i-1} \mathbf{q}_j\mathbf{q}_j^*)\mathbf{a}_i\|}
\end{aligned} \tag{45}$$

This gave rise to three algorithms

- Classical Gram-Schmidt
- Modified Gram-Schmidt
- Iterative Gram-Schmidt

Together with their operational count.

Definition 0.11.4. *Let $\mathbf{v} \in \mathbb{K}^n$ be a normal vector defining a hyperplane. The transformation*

$$f_{\mathrm{H}} : \mathbb{K}^n \to \mathbb{K}^n \ : \ x \mapsto x - 2\langle x, v \rangle v$$

is the Householder transformation about the hyperplane defined by the normal vector $\mathbf{v} \in \mathbb{K}^n$.

Proposition 0.11.2. *Let $\mathbf{v} \in \mathbb{K}^n$ be a normal vector defining a hyperplane and f_{H} be the Householder transformation about the hyperplane defined by the normal vector $\mathbf{v} \in \mathbb{K}^n$. Then f_{H} is a linear map and its matrix representation is*

$$\mathbf{P}_{\mathbf{v}} = \mathbf{I} - 2\mathbf{v}\mathbf{v}^*$$

Proposition 0.11.3. *Let $\mathbf{v} \in \mathbb{K}^n$ be a normal vector defining a hyperplane and f_{H} be the Householder transformation about the hyperplane defined by the normal vector $\mathbf{v} \in \mathbb{K}^n$. The householder matrix $\mathbf{P}_{\mathbf{v}}$ fulfills:*

i)	Hermitian	$(\mathbf{P}_{\mathbf{v}} = \mathbf{P}_{\mathbf{v}}^*)$	iv)	$\mathbf{P}_{\mathbf{v}}$ has eigenvalues ± 1
ii)	Unitary	$(\mathbf{P}_{\mathbf{v}}^{-1} = \mathbf{P}_{\mathbf{v}}^*)$	v)	$\det(\mathbf{P}_{\mathbf{v}}) = -1$
iii)	Involutory	$(\mathbf{P}_{\mathbf{v}}^{-1} = \mathbf{P}_{\mathbf{v}})$		

Proof. First note that

$$\mathbf{P}_{\mathbf{v}}^* = (\mathbf{I} - 2\mathbf{v}\mathbf{v}^*)^* = \mathbf{I} - 2\mathbf{v}\mathbf{v}^* = \mathbf{P}_{\mathbf{v}} \tag{46}$$

which shows i). Next, we consider

$$\mathbf{P}_{\mathbf{v}}^2 = (\mathbf{I} - 2\mathbf{v}\mathbf{v}^*)(\mathbf{I} - 2\mathbf{v}\mathbf{v}^*) = \mathbf{I} - 4\mathbf{v}\mathbf{v}^* + 4\mathbf{v}\mathbf{v}^* = \mathbf{I} \tag{47}$$

showing that $\mathbf{P}_{\mathbf{v}}$ is involutory. This in turn yields that $\mathbf{P}_{\mathbf{v}}$ is unitary, since

$$\mathbf{P}_{\mathbf{v}}^{-1} = \mathbf{P}_{\mathbf{v}} = \mathbf{P}_{\mathbf{v}}^*. \tag{48}$$

Note that for $\mathbf{u} \perp \mathbf{v}$ we have $\mathbf{P}_{\mathbf{v}}\mathbf{u} = \mathbf{u}$. Since there are $n-1$ linearly independent vectors $\mathbf{u} \in \mathbb{K}^n$ fulfilling $\mathbf{u} \perp \mathbf{v}$, the eigenspace of $\mathbf{P}_{\mathbf{v}}$ corresponding to the eigenvalue $\lambda = 1$ is $n-1$ dimensional. Moreover $\mathbf{P}_{\mathbf{v}}\mathbf{v} = -\mathbf{v}$, showing iv). By iv), we know that $\mathbf{P}_{\mathbf{v}}$ is diagonalizable with $n-1$ eigenvalues $\lambda_1 = 1$ and one eigenvalue $\lambda_1 = -1$. Aplying the determinant multiplication Theorem we have

$$\det(\mathbf{P}_{\mathbf{v}}) = \det \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \dots & 0 & -1 \end{bmatrix} = (-1)\, 1^{n-1} = -1 \tag{49}$$

$\square$

The most important application:

Householder QR

We also did its operational count and argued how to keep it low:

Work with Householder vectors

Definition 0.11.5. *Let $i, j \in \llbracket m \rrbracket$ and $\theta \in [0, 2\pi)$. A matrix $\mathbf{G}(i, j, \theta) \in \mathbb{K}^{m \times m}$ defined through*

$$[\mathbf{G}(i,j,\theta)]_{l,m} = \begin{cases} 1 & \text{, if } l = m, \text{ and } l \neq i, j \\ \cos(\theta) & \text{, if } l = m = i, j \\ \sin(\theta) & \text{, if } l = i, \text{ and } m = j \\ -\sin(\theta) & \text{, if } l = j, \text{ and } m = i \\ 0 & \text{, else.} \end{cases} \tag{50}$$

is called Givens rotation around θ in the i-j-plane.

Proposition 0.11.4. *Givens rotations are orthogonal matrices, i.e., $\mathbf{G}^\top = \mathbf{G}^{-1}$.*

Remark 0.11.3. *Givens rotations indeed rotate in the i-j-plane. Consider*

$$\mathbf{G}(i,j,\theta)\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_{i-1} \\ cx_i - sx_j \\ x_{i+1} \\ \vdots \\ cx_j + sx_i \\ x_{j+1} \\ \vdots \\ x_m \end{bmatrix} \Bigg] = \begin{bmatrix} c & -s \\ s & c \end{bmatrix} \begin{bmatrix} x_i \\ x_j \end{bmatrix} \tag{51}$$

substituting c and s with $\cos(\theta)$ and $\sin(\theta)$, respectively, we see that this corresponds to a (counter-clockwise) rotation through an angle θ in the i-j-plane.

We designed an algorithm that uses Givens rotations to compute a QR factorization and discussed the operational count, and how to keep it low:

Track only the Givens angles.