Appendix A.

Additional lemmas

A.1. Matrix properties

A.1.1. Characterization of uniform positive definite matrices

Lemma A.1.1 (Rayleigh quotient). *Let $A \in \mathbb{C}^{n\times n}$ be a hermitian matrix, with ordered eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ (ignoring their multiplicities). Then for all $x \in \mathbb{C}^n\backslash\{0\}$ we get*

$$\lambda_1 \leq \frac{x^\mathsf{T} A x}{\|x\|_2^2} \leq \lambda_n. \tag{A.1}$$

Proof. The spectral theorem [Fis13, p. 307] yields, that we have an orthonormal basis $\{v_i\}_{i=1}^n$ of eigenvectors $v_i \in \mathbb{C}^n\backslash\{0\}$ with $Av_i = \lambda_i v_i$ for $i = 1, \ldots, n$. Using the Bessel equation [Fis13, p. 300], we obtain for arbitrary $x \in \mathbb{C}^n\backslash\{0\}$ with $x = \sum_{i=1}^n \langle x, v_j\rangle v_j$, that

$$\|x\|_2^2 \overset{\text{Bessel}}{=} \sum_{i=1}^n |\langle x, v_i\rangle|^2$$

holds. Therefore, our claim follows as an easy consequence, since we have

$$\begin{aligned} \frac{x^\mathsf{T} A x}{\|x\|_2^2} &= \frac{\langle \sum_{i=1}^n \lambda_i \langle x, v_j\rangle v_j, x\rangle}{\|x\|_2^2} \\ &\leq \lambda_n \frac{\sum_{i=1}^n |\langle x, v_i\rangle|^2}{\|x\|_2^2} = \lambda_n. \end{aligned}$$

The other inequality follows analogously. □

Corollary A.1.2 (Courant–Fischer–Weyl min-max principle)**.** *Let $A \in \mathbb{C}^{n\times n}$ be a hermitian matrix, with ordered eigenvalues $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ (ignoring their multiplicities). Then we can estimate the minimal and maximal eigenvalue of A via*

$$\mu_{\min}(A) = \min_{x\in\mathbb{C}^n\setminus\{0\}} \frac{x^\mathsf{T} Ax}{\|x\|_2^2}$$
$$\mu_{\max}(A) = \max_{x\in\mathbb{C}^n\setminus\{0\}} \frac{x^\mathsf{T} Ax}{\|x\|_2^2}.$$

Proof. The result is a direct consequence of the proof of Lemma A.1.1. We just need to consider the eigenpairs (λ_1, v_1) and (λ_n, v_n), follow the proof with $x := v_1$, respectively $x := v_n$, and obtain the desired equalities. □

Corollary A.1.3 (Hermitian positive definite matrices are uniform positive definite)**.** *Let $A \in \mathbb{C}^{n\times n}$ be a hermitian positive definite matrix, then A is uniform positive definite, meaning that there exists a $C \in \mathbb{R}^+$, such that for every $d \in \mathbb{C}^n$ the inequality*

$$d^\mathsf{T} Ad \ge C \cdot \|d\|_2^2$$

holds. In particular, we can choose $C = \lambda_{\min}(A) \in \mathbb{R}^+$.

Proof. Using Lemma A.1.1, we observe for $d \in \mathbb{C}^n$

$$d^\mathsf{T} Ad \ge \lambda_{\min}(A) \cdot \|d\|_2^2.$$

Since A is positive definite, we have $\lambda_{\min}(A) > 0$. By defining $C := \lambda_{\min}(A)$, we obtain the desired result. □

A.1.2. Convergence analysis of the Jacobi method

Definition A.1.4 (Spectral radius of a matrix)**.** The **spectral radius** of a matrix $A \in \mathbb{R}^{n\times n}$ is defined as

$$\rho(A) := \max_{j=1,\ldots,n} |\lambda_j(A)|,$$

where $\lambda_j(A) \in \mathbb{C}$ denotes the j-th eigenvalue of the matrix A (ignoring multiplicities).

Lemma A.1.5 (Basic properties of the spectral radius)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a matrix. For all induced matrix norms, $\rho(A) \le \|A\|$ holds. If A is additionally symmetric and $\|\cdot\| = \|\cdot\|_2$ is the 2-norm, we have $\rho(A) = \|A\|_2$.*

Proof. First, let $A \in \mathbb{R}^{n \times n}$ be an arbitrary matrix and (λ, v) be an eigenpair of A. Using the relation $Av = \lambda v$, we obtain

$$|\lambda| \|v\| = \|\lambda v\| = \|Av\| \leq \|A\| \|v\|,$$

and since $\|v\| \neq 0$, we have $|\lambda| \leq \|A\|$. Since this holds for any eigenvalue, the first claim follows.

For the second claim we assume that A is symmetric. By the definition of the 2-Norm we observe

$$\|A\|_2^2 = \lambda_{\max}(A^\mathsf{T} A) = \lambda_{\max}(A^2) = \max |\lambda(A)|^2 = \rho(A)^2.$$

The claim follows. □

Theorem A.1.6 (Convergence of stationary methods)**.** *Let $A \in \mathbb{R}^{n \times n}$ be a regular matrix, $A = M - N$ with M regular and $f \in \mathbb{R}^n$. The stationary iterative method*

$$Mx_{k+1} = Nx_k + f$$

converges for any initial vector $x_0 \in \mathbb{R}^n$ to the solution x of the linear system $Ax = f$ if and only if $\rho(M^{-1}N) < 1$.

Proof. See [CG20, Theorem 3]. □

Corollary A.1.7 (Convergence of the Jacobi method for strictly diagonally dominant matrices)**.** *If the matrix $A \in \mathbb{R}^{n \times n}$ is strictly diagonally dominant, i.e.*

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|, \qquad \text{for } i = 1, \ldots, n, \tag{A.2}$$

then the Jacobi method, i.e. the iterative method induced by the splitting $M = \operatorname{diag}(A)$ and $N = \operatorname{diag}(A) - A$, converges.

Proof. The condition (A.2) allows us to estimate

$$\left\|M^{-1}N\right\|_\infty = \max_{i=1,\ldots,n} \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}| < 1.$$

Using Lemma A.1.5, we observe

$$\rho(M^{-1}N) \leq \left\|M^{-1}N\right\|_{\infty} < 1,$$

which, together with Theorem A.1.6, yields the desired result. □

A.2. Probability theory

A.2.1. Convergence in probability and Markov's inequality

Definition A.2.1 (Convergence in probability)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a sequence of real-valued random variables $\{X_n\}_{n\in\mathbb{N}} : \Omega \to \mathbb{R}$. We say $\{X_n\}_{n\in\mathbb{N}}$ converges **in probability** to a real-valued random variable $X : \Omega \to \mathbb{R}$, if for every $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| \geq \varepsilon\right) = 0$$

holds. We write $X_n \overset{\mathbb{P}}{\to} X$.

Lemma A.2.2 (Markov's inequality)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $X : \Omega \to \mathbb{R}$ a real-valued random variable and $h : \mathbb{R} \to [0, \infty)$ a monotonically increasing function. Let $\varepsilon \geq 0$ with $h(\varepsilon) > 0$, then*

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}\left[h(X)\right]}{h(\varepsilon)}.$$

Proof. As a monotone function, h is $\mathcal{B}(\mathbb{R})$-measurable. Hence, we have

$$\mathbb{P}(X \geq \varepsilon) = \int_{\Omega} \chi_{\{X\geq\varepsilon\}} \,\mathrm{d}\mathbb{P} \leq \int_{\Omega} \chi_{\{X\geq\varepsilon\}} \frac{h(X)}{h(\varepsilon)} \,\mathrm{d}\mathbb{P} \leq \frac{\mathbb{E}\left[h(X)\right]}{h(\varepsilon)}.$$

□

A.3. Functional analysis

A.3.1. The lemmas of Fréchet-Riesz, Lax-Milgram and Friedrich

Lemma A.3.1 (Fréchet-Riesz). *Consider a Hilbert space $(H, \langle \cdot, \cdot \rangle_V)$ with its dual space H'. For every $u' \in H'$, there exists a unique $u \in H$ with*

$$\langle u', v \rangle_{H' \times H} = \langle u, v \rangle_H \ \forall v \in H$$

and $\|u'\|_{H'} = \|u\|_H$.

Proof. See [Wer07, Theorem V.3.6, p. 226]. □

Lemma A.3.2 (Lax-Milgaram). *Let H be a Hilbert space over a field $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and $B : H \times H \to \mathbb{K}$ a continuous bilinear function. Further, assume having a constant $p > 0$ with*

$$|B(u,u)| \geq p \|u\|^2 \ \ \forall u \in H.$$

Then there exists for every $F \in H'$ a unique $u \in H$, such that

$$F(v) = B(u,v) \ \forall v \in H$$

holds.

Proof. See [Den21, Thm. 6.6]. □

Lemma A.3.3 (Friedrichs inequality). *Let $G \subset \mathbb{R}^n$ be a bounded Lipschitz domain of $\mathbb{R}^n$ and $\mathrm{d}(G) := \sup_{x,y \in G} \|x - y\|_{\mathbb{R}^n}$ the diameter of the domain. Then for every $u \in H_0^1(G)$ the inequality*

$$\int_G u(x) \, \mathrm{d}x \leq \mathrm{d}(G) \int_G |\nabla u(x)|^2 \ \mathrm{d}x$$

holds.

Proof. See [Rek12, Thm. 18.1, p. 118]. □

Bibliography

[Aln+15] Martin S. Alnæs et al. "The FEniCS Project Version 1.5". In: *Archive of Numerical Software* 3.100 (2015). DOI: `10.11588/ans.2015.100.20553`.

[BCN18] Léon Bottou, Frank E. Curtis, and Jorge Nocedal. "Optimization Methods for Large-Scale Machine Learning". In: *SIAM Review* 60.2 (2018), pp. 223–311. DOI: `10.1137/16M1080173`. eprint: `https://doi.org/10.1137/16M1080173`. URL: `https://doi.org/10.1137/16M1080173`.

[CG20] Gabriele Ciaramella and Martin J. Gander. *Iterative Methods and Preconditioners for Systems of Linear Equations*. 2020.

[CS10] Saifon Chaturantabut and Danny C. Sorensen. "Nonlinear Model Reduction via Discrete Empirical Interpolation". In: *SIAM Journal on Scientific Computing* 32.5 (2010), pp. 2737–2764.

[CS12] Saifon Chaturantabut and Danny C. Sorensen. "A state space error estimate for POD-DEIM nonlinear model reduction". In: *SIAM Journal on Numerical Analysis* 50.1 (2012), pp. 46–63.

[Den21] Robert Denk. "Theorie partieller Differentialgleichungen – Vorlesungsskript". Konstanz, 2021.

[DR08] Wolfgang Dahmen and Arnold Reusken. *Numerik für Ingenieure und Naturwissenschaftler*. 2nd ed. Berlin Heidelberg New York: Springer Spektrum, 2008. ISBN: 978-3-540-76493-9.

[Dzi10] Gerhard Dziuk. *Theorie und Numerik partieller Differentialgleichungen*. Berlin: Walter de Gruyter, 2010. ISBN: 978-3-110-14843-5.

[Fis13] Gerd Fischer. *Lineare Algebra - Eine Einführung für Studienanfänger*. 18th ed. Berlin Heidelberg New York: Springer Spektrum, 2013. ISBN: 978-3-658-03945-5.

[Jun21] Michael Junk. "Numerik partieller Differentialgleichungen – Vorlesungsskript". Konstanz, 2021.

[Kle13] Achim Klenke. *Wahrscheinlichkeitstheorie*. 3rd ed. Berlin Heidelberg New York: Springer Spektrum, 2013. ISBN: 978-3-642-36017-6.

[LGY20] Yanli Liu, Yuan Gao, and Wotao Yin. *An Improved Analysis of Stochastic Gradient Descent with Momentum*. 2020. arXiv: 2007.07989 [math.OC].

[Pas+19] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[Rek12] Karel Rektorys. *Variational Methods in Mathematics, Science and Engineering*. Berlin Heidelberg: Springer Science & Business Media, 2012. ISBN: 978-9-401-16450-4.

[Wer07] Dirk Werner. *Funktionalanalysis*. 6th ed. Berlin Heidelberg New York: Springer-Verlag, 2007. ISBN: 978-3-540-72536-7.

[WH21] Florian Wolf and Franz Herbst. *A machine learning approach to predict a vehicles velocity using dashcam video footage*. Tech. rep. Konstanz: University Konstanz, Department of Computer and Information Science, 2021.