Chapter 5
Conclusion

This thesis provided insights in the topics of consistency, sparsity and robustness of learning algorithms. Thematically, the thesis is split into two main chapters, dealing with semi-supervised and supervised learning. We summarize both chapters below and respectively provide an outlook and possible future work.

Consistency of SSL on Sparse Graphs We considered the infinite data limit in the semi-supervised setting, focusing on the Lipschitz learning task. For all our results we assumed a mild scaling conditions which allow for very sparse graphs. We were first able to show Γ-convergence in the variational setting. We then proved convergence rates for AMLEs via a homogenization strategy, where we the proof relied on the comparison with cones principle. The key insight we obtained was, that a rate for graph distance functions implies a rate for graph AMLEs. This observation was already employed in our follow-up work in [LIP-III], where convergence rate at a even smaller scale was shown. We also conducted experiments to validate our theoretical framework in practice. Here, we observed better results than we were able to prove.

While our framework allows the graph to be very sparse, we are restricted to the case of ball graphs. Here it would be interesting to see, how our results can be transferred to the knn setting as in [CT22]. Furthermore, we only considered the standard Lipschitz learning task, that tends to forget the data distribution, which is undesirable for many learning tasks. In [Cal19] a modification was proposed that makes the problem sensitive to the distribution. The open problem that arises here, is how to adapt the technique in [LIP-II] to show convergence for the modified problem.

Robust and Sparse SL In the supervised setting, we considered input-robustness w.r.t. adversarial perturbations and resolution changes. In the first case we proposed a defense mechanism to train stable neural networks, based on Lipschitz regularization. We provided analytical results and numerical experiments which suggest that the strategy allows to learn robust networks. The strength of this approach depends on how well the discrete Lipschitz constant approximates the true one. An interesting future direction would be to explore different methods to determine the discrete set of Lipschitz pairs.

E.g. one could additionally learn a generative model that outputs these pair in a faster and probably even more expressive way than the gradient descent scheme, employed right now.

For the multi-resolution setting we analyzed the role of Fourier neural operators. We first showed functional analytic properties of neural layers acting on L^p spaces. We then established the relation of disctretized Fourier layers to standard convolution operations, both from a theoretical and a practical side. We also highlighted the importance of the trigonometric interpolation in this context. We conducted numerical tests that suggest the resolution equivariant behavior of FNO layers. Connected to the previous topic it would be interesting to study the adversarial robustness of FNOs. First numerical tests in [Kab22] hint that the vulnerabilities are even worse than in the standard case.

We considered the question on how to enforce sparsity in neural network weights. Here, we employed a stochastic variant of Bregman iterations, which allowed to train networks that are very sparse throughout the whole optimization. We provided theoretic convergence guarantees, where the main novelty was the stochasticity introduced into the iteration. We also showed the numerical efficiency of the method, by training sparse neural networks performing similarly to their dense counterparts. One open problem is to weaken the convexity assumption and instead prove the convergence result, by employing a Kurdyka–Łojasiewicz type inequality as in [Ben+21]. Concerning the neural architecture search via sparsity as proposed in our work, a further interesting task would be to learn the U-Net type architecture of [RFB15].

Bibliography

Books

[Ros+62] F. Rosenblatt et al. *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Vol. 55. Spartan books Washington, DC, 1962.

[AF03] R. A. Adams and J. J. Fournier. *Sobolev spaces*. Elsevier, 2003.

[GV13] G. H. Golub and C. F. Van Loan. *Matrix computations*. JHU press, 2013.

[Lip77] R. Lipschitz. *Lehrbuch der analysis*. Vol. 1. M. Cohen & Sohn (F. Cohen), 1877.

[Bre+84] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Clasificaion and regression trees*. CRC Press, New York, 1984.

[Zhu05] X. Zhu. *Semi-supervised learning with graphs*. Carnegie Mellon University, 2005.

[Lin17] P. Lindqvist. *Notes on the p-Laplace equation*. 161. University of Jyväskylä, 2017.

[Eva18] L. Evans. *Measure theory and fine properties of functions*. Routledge, 2018.

[BB11] H. Brezis and H. Brézis. *Functional analysis, Sobolev spaces and partial differential equations*. Vol. 2. 3. Springer, 2011.

[Sch69] J. T. Schwartz. *Nonlinear Functional Analysis -*. Boca Raton, Fla: CRC Press, 1969.

[Lin16] P. Lindqvist. *Notes on the infinity Laplace equation*. Springer, 2016.

[Bra02] A. Braides. *Gamma-convergence for Beginners*. Vol. 22. Oxford University Press, Oxford, 2002.

[Dal12] G. Dal Maso. *An introduction to* Γ*-convergence*. Vol. 8. Springer Science & Business Media, 2012.

[Hau14] F. Hausdorff. *Grundzüge der Mengenlehre*. Viet, Leipzig, 1914.

[DS88] N. Dunford and J. T. Schwartz. *Linear operators, part 1: general theory*. Vol. 10. John Wiley & Sons, 1988.

[COR98] G. Cybenko, D. P. O'Leary, and J. Rissanen. *The mathematics of information coding, extraction and distribution*. Vol. 107. Springer Science & Business Media, 1998.

[Dac07] B. Dacorogna. *Direct methods in the calculus of variations*. Vol. 78. Springer Science & Business Media, 2007.

[SB14] S. Shalev-Shwartz and S. Ben-David. *Understanding machine learning: From theory to algorithms.* Cambridge university press, 2014.

[VD95] G. Van Rossum and F. L. Drake Jr. *Python reference manual.* Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[Roc97] R. Rockafellar. *Convex analysis.* Princeton, N.J: Princeton University Press, 1997.

[BC11] H. Bauschke and P. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces.* New York: Springer, 2011.

[Eul24] L. Euler. *Institutionum calculi integralis.* Vol. 1. impensis Academiae imperialis scientiarum, 1824.

[BV04] S. P. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge university press, 2004.

[GBC16] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning.* http://www.deeplearningbook.org. MIT Press, 2016.

[Ral81] L. B. Rall. *Automatic differentiation: Techniques and applications.* Springer, 1981.

[GW87] R. C. Gonzales and P. Wintz. *Digital image processing.* Addison-Wesley Longman Publishing Co., Inc., 1987.

[Trö10] F. Tröltzsch. *Optimal Control of Partial Differential Equations: Theory, Methods, and Applications.* Vol. 112. Graduate Studies in Mathematics. American Mathematical Society, Providence, Rhode Island, 2010.

[Gra14] L. Grafakos. *Classical Fourier Analysis.* 3rd ed. Graduate Texts in Mathematics. Springer, New York, NY, 2014.

[AP93] A. Ambrosetti and G. Prodi. *A Primer of Nonlinear Analysis.* Cambridge University Press, 1993.

Articles

[LIP-I] T. Roith and L. Bungert. “Continuum limit of Lipschitz learning on graphs.” In: *Foundations of Computational Mathematics* (2022), pp. 1–39.

[LIP-II] L. Bungert, J. Calder, and T. Roith. “Uniform convergence rates for Lipschitz learning on graphs.” In: *IMA Journal of Numerical Analysis* (Sept. 2022). DOI: 10.1093/imanum/drac048.

[BREG-I] L. Bungert, T. Roith, D. Tenbrinck, and M. Burger. “A Bregman learning framework for sparse neural networks.” In: *Journal of Machine Learning Research* 23.192 (2022), pp. 1–43.

[Sam59] A. L. Samuel. “Some studies in machine learning using the game of checkers.” In: *IBM Journal of research and development* 3.3 (1959), pp. 210–229.

[Ros58] F. Rosenblatt. "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6 (1958), p. 386.

[Kel60] H. J. Kelley. "Gradient theory of optimal flight paths." In: *Ars Journal* 30.10 (1960), pp. 947–954.

[RHW86] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning representations by back-propagating errors." In: *nature* 323.6088 (1986), pp. 533–536.

[Sch22] J. Schmidhuber. "Annotated history of modern AI and Deep learning." In: *arXiv preprint arXiv:2212.11279* (2022).

[Cha+21] J. Chai, H. Zeng, A. Li, and E. W. Ngai. "Deep learning in computer vision: A critical review of emerging techniques and application scenarios." In: *Machine Learning with Applications* 6 (2021), p. 100134.

[Khu+23] D. Khurana, A. Koli, K. Khatter, and S. Singh. "Natural language processing: State of the art, current trends and challenges." In: *Multimedia tools and applications* 82.3 (2023), pp. 3713–3744.

[She+22] M. Shehab, L. Abualigah, Q. Shambour, M. A. Abu-Hashem, M. K. Y. Shambour, A. I. Alsalibi, and A. H. Gandomi. "Machine learning in medical applications: A review of state-of-the-art methods." In: *Computers in Biology and Medicine* 145 (2022), p. 105458.

[GSS14] I. J. Goodfellow, J. Shlens, and C. Szegedy. "Explaining and harnessing adversarial examples." In: *arXiv preprint arXiv:1412.6572* (2014).

[NSZ09] B. Nadler, N. Srebro, and X. Zhou. "Statistical analysis of semi-supervised learning: The limit of infinite unlabelled data." In: *Advances in neural information processing systems* 22 (2009).

[Hoe+21] T. Hoefler, D. Alistarh, T. Ben-Nun, N. Dryden, and A. Peste. "Sparsity in Deep Learning: Pruning and growth for efficient inference and training in neural networks." In: *J. Mach. Learn. Res.* 22.241 (2021), pp. 1–124.

[Lan52] C. Lanczos. "Solution of systems of linear equations by minimized iterations." In: *J. Res. Nat. Bur. Standards* 49.1 (1952), pp. 33–53.

[Li+20] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. "Fourier neural operator for parametric partial differential equations." In: *arXiv preprint arXiv:2010.08895* (2020).

[Fuk80] K. Fukushima. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position." In: *Biological cybernetics* 36.4 (1980), pp. 193–202.

[Osh+05] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. "An iterative regularization method for total variation-based image restoration." In: *Multiscale Modeling & Simulation* 4.2 (2005), pp. 460–489.

[Bol68] L. Boltzmann. “Studien über das Gleichgewicht der lebenden Kraft.” In: *Wissenschafiliche Abhandlungen* 1 (1868), pp. 49–96.

[ST14] A. Subramanya and P. P. Talukdar. “Graph-based semi-supervised learning.” In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* 8.4 (2014), pp. 1–125.

[Ste+56] H. Steinhaus et al. “Sur la division des corps matériels en parties.” In: *Bull. Acad. Polon. Sci* 1.804 (1956), p. 801.

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. “Maximum likelihood from incomplete data via the EM algorithm.” In: *Journal of the royal statistical society: series B (methodological)* 39.1 (1977), pp. 1–22.

[GS15] N. García Trillos and D. Slepčev. “Continuum Limit of Total Variation on Point Clouds.” In: *Archive for Rational Mechanics and Analysis* 220.1 (2015), pp. 193–241. DOI: 10.1007/s00205-015-0929-z.

[SB09] A. Szlam and X. Bresson. “A total variation-based graph clustering algorithm for cheeger ratio cuts.” In: *UCLA Cam report* (2009), pp. 09–68.

[Gar+16] N. García Trillos, D. Slepčev, J. Von Brecht, T. Laurent, and X. Bresson. “Consistency of Cheeger and ratio graph cuts.” In: *The Journal of Machine Learning Research* 17.1 (2016), pp. 6268–6313.

[GMT22] N. García Trillos, R. Murray, and M. Thorpe. “From graph cuts to isoperimetric inequalities: Convergence rates of Cheeger cuts on data clouds.” In: *Archive for Rational Mechanics and Analysis* 244.3 (2022), pp. 541–598.

[GS18] N. García Trillos and D. Slepčev. “A variational approach to the consistency of spectral clustering.” In: *Applied and Computational Harmonic Analysis* 45.2 (2018), pp. 239–281.

[THH21] N. G. Trillos, F. Hoffmann, and B. Hosseini. “Geometric structure of graph Laplacian embeddings.” In: *The Journal of Machine Learning Research* 22.1 (2021), pp. 2934–2988.

[Hof+22] F. Hoffmann, B. Hosseini, A. A. Oberai, and A. M. Stuart. “Spectral analysis of weighted Laplacians arising in data clustering.” In: *Applied and Computational Harmonic Analysis* 56 (2022), pp. 189–249.

[CV95] C. Cortes and V. Vapnik. “Support-vector networks.” In: *Machine learning* 20 (1995), pp. 273–297.

[MS63] J. N. Morgan and J. A. Sonquist. “Problems in the analysis of survey data, and a proposal.” In: *Journal of the American statistical association* 58.302 (1963), pp. 415–434.

[MP69] M. Minsky and S. Papert. “An introduction to computational geometry.” In: *Cambridge tiass., HIT* 479 (1969), p. 480.

[Sch15] J. Schmidhuber. “Deep learning in neural networks: An overview.” In: *Neural Networks* 61 (2015), pp. 85–117. DOI: https://doi.org/10.1016/j.neunet.2014.09.003.

[ST19] D. Slepcev and M. Thorpe. “Analysis of p-Laplacian regularization in semisupervised learning.” In: *SIAM Journal on Mathematical Analysis* 51.3 (2019), pp. 2085–2120.

[Cal19] J. Calder. “Consistency of Lipschitz learning with infinite unlabeled data and finite labeled data.” In: *SIAM Journal on Mathematics of Data Science* 1.4 (2019), pp. 780–812.

[FCL19] M. Flores, J. Calder, and G. Lerman. “Algorithms for Lp-based semi-supervised learning on graphs.” In: *arXiv preprint arXiv:1901.05031* (2019).

[CT22] J. Calder and N. G. Trillos. “Improved spectral convergence rates for graph Laplacians on ε-graphs and k-NN graphs.” In: *Applied and Computational Harmonic Analysis* 60 (2022), pp. 123–175.

[ACJ04] G. Aronsson, M. Crandall, and P. Juutinen. “A tour of the theory of absolutely minimizing functions.” In: *Bulletin of the American mathematical society* 41.4 (2004), pp. 439–505.

[ETT15] A. Elmoataz, M. Toutain, and D. Tenbrinck. “On the p-Laplacian and ∞-Laplacian on graphs with applications in image and data processing.” In: *SIAM Journal on Imaging Sciences* 8.4 (2015), pp. 2412–2451.

[AL11] M. Alamgir and U. Luxburg. “Phase transition in the family of p-resistances.” In: *Advances in neural information processing systems* 24 (2011).

[VBB08] U. Von Luxburg, M. Belkin, and O. Bousquet. “Consistency of spectral clustering.” In: *The Annals of Statistics* (2008), pp. 555–586.

[GK06] E. Giné and V. Koltchinskii. “Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results.” In: *Lecture Notes-Monograph Series* (2006), pp. 238–259.

[Hof+20] F. Hoffmann, B. Hosseini, Z. Ren, and A. M. Stuart. “Consistency of semi-supervised learning algorithms on graphs: Probit and one-hot methods.” In: *The Journal of Machine Learning Research* 21.1 (2020), pp. 7549–7603.

[Dun+20] M. M. Dunlop, D. Slepčev, A. M. Stuart, and M. Thorpe. “Large data and zero noise limits of graph-based semi-supervised learning algorithms.” In: *Applied and Computational Harmonic Analysis* 49.2 (2020), pp. 655–697.

[CS20] J. Calder and D. Slepčev. “Properly-weighted graph Laplacian for semi-supervised learning.” In: *Applied mathematics & optimization* 82 (2020), pp. 1111–1159.

[vLB04] U. von Luxburg and O. Bousquet. “Distance-Based Classification with Lipschitz Functions.” In: *J. Mach. Learn. Res.* 5.Jun (2004), pp. 669–695.

[Jen93] R. Jensen. “Uniqueness of Lipschitz extensions: minimizing the sup norm of the gradient.” In: *Archive for Rational Mechanics and Analysis* 123 (1993), pp. 51–74.

[Kir34] M. Kirszbraun. "Über die zusammenziehende und Lipschitzsche Transformationen." In: *Fundamenta Mathematicae* 22 (1934), pp. 77–108. DOI: 10.4064/fm-22-1-77-108.

[Whi92] H. Whitney. "Analytic extensions of differentiable functions defined in closed sets." In: *Hassler Whitney Collected Papers* (1992), pp. 228–254.

[McS34] E. J. McShane. "Extension of range of functions." In: (1934).

[Aro67] G. Aronsson. "Extension of functions satisfying Lipschitz conditions." In: *Arkiv för Matematik* 6.6 (1967), pp. 551–561.

[Per23] O. Perron. "Eine neue Behandlung der ersten Randwertaufgabe für Δu= 0." In: *Mathematische Zeitschrift* 18 (1923), pp. 42–54.

[Aro68] G. Aronsson. "On the partial differential equation $u_x^2 u_{xx} + 2u_x u_y u_{xy} + u_y^2 u_{yy} = 0$." In: *Arkiv för matematik* 7 (1968), pp. 395–425.

[Yu06] Y. Yu. "A remark on C2 infinity-harmonic functions." In: (2006).

[BDM89] T. Bhattacharya, E. DiBenedetto, and J. Manfredi. "Limits as $p \to \infty$ of $Delta_p u_p = f$ and related extremal problems." In: *Rend. Sem. Mat. Univ. Politec. Torino* 47 (1989), pp. 15–68.

[Bur48] J. M. Burgers. "A mathematical model illustrating the theory of turbulence." In: *Advances in applied mechanics* 1 (1948), pp. 171–199.

[ASS11] S. Armstrong, C. Smart, and S. Somersille. "An infinity Laplace equation with gradient term and mixed boundary conditions." In: *Proceedings of the American Mathematical Society* 139.5 (2011), pp. 1763–1776.

[AS10] S. N. Armstrong and C. K. Smart. "An easy proof of Jensen's theorem on the uniqueness of infinity harmonic functions." In: *Calculus of Variations and Partial Differential Equations* 37 (2010), pp. 381–384.

[JS06] P. Juutinen and N. Shanmugalingam. "Equivalence of AMLE, strong AMLE, and comparison with cones in metric measure spaces." In: *Mathematische Nachrichten* 279.9-10 (2006), pp. 1083–1098.

[Per+09] Y. Peres, O. Schramm, S. Sheffield, and D. Wilson. "Tug-of-war and the infinity Laplacian." In: *Journal of the American Mathematical Society* 22.1 (2009), pp. 167–210.

[NS12] A. Naor and S. Sheffield. "Absolutely minimal Lipschitz extension of tree-valued mappings." In: *Mathematische Annalen* 354 (2012), pp. 1049–1078.

[Kur22] C. Kuratowski. "Sur l'opération A de l'analysis situs." In: *Fundamenta Mathematicae* 3.1 (1922), pp. 182–199.

[ČFK66] E. Čech, Z. Frolík, and M. Katětov. "Topological spaces." In: (1966).

[DF75] E. De Giorgi and T. Franzoni. "Su un tipo di convergenza variazionale." In: *Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti* 58.6 (1975), pp. 842–850.

[Mod77] L. Modica. “Un esempio di Γ-convergenza.” In: *Boll. Un. Mat. Ital. B* 14 (1977), pp. 285–299.

[CGL10] A. Chambolle, A. Giacomini, and L. Lussardi. “Continuous limits of discrete perimeters.” In: *ESAIM: Mathematical Modelling and Numerical Analysis* 44.2 (2010), pp. 207–230.

[BY12] A. Braides and N. K. Yip. “A quantitative description of mesh dependence for the discretization of singularly perturbed nonconvex problems.” In: *SIAM Journal on Numerical Analysis* 50.4 (2012), pp. 1883–1898.

[VB+12] Y. Van Gennip, A. L. Bertozzi, et al. “Γ-convergence of graph Ginzburg-Landau functionals.” In: *Adv. Differential Equations* 17.11-12 (2012), pp. 1115–1180.

[Fré06] M. M. Fréchet. “Sur quelques points du calcul fonctionnel.” In: *Rendiconti del Circolo Matematico di Palermo (1884-1940)* 22.1 (1906), pp. 1–72.

[Pen99a] M. D. Penrose. “A strong law for the longest edge of the minimal spanning tree.” In: *The Annals of Probability* 27.1 (1999), pp. 246–260.

[Pen99b] M. D. Penrose. “A strong law for the largest nearest-neighbour link between random points.” In: *Journal of the london mathematical society* 60.3 (1999), pp. 951–960.

[TS15] N. G. Trillos and D. Slepčev. “On the rate of convergence of empirical measures in ∞-transportation distance.” In: *Canadian Journal of Mathematics* 67.6 (2015), pp. 1358–1383.

[San15] F. Santambrogio. “Optimal transport for applied mathematicians.” In: *Birkäuser, NY* 55.58-63 (2015), p. 94.

[BKB20] L. Bungert, Y. Korolev, and M. Burger. “Structural analysis of an L-infinity variational problem and relations to distance functions.” In: *Pure and Applied Analysis* 2.3 (2020), pp. 703–738. DOI: 10.2140/paa.2020.2.703.

[Goo52] I. J. Good. “Rational decisions.” In: *Journal of the Royal Statistical Society: Series B (Methodological)* 14.1 (1952), pp. 107–114.

[SGS15] R. K. Srivastava, K. Greff, and J. Schmidhuber. “Highway networks.” In: *arXiv preprint arXiv:1505.00387* (2015).

[BREG-II] L. Bungert, T. Roith, D. Tenbrinck, and M. Burger. “Neural Architecture Search via Bregman Iterations.” In: (2021). arXiv: 2106.02479 [cs.LG].

[Rie22] K. Riedl. “Leveraging memory effects and gradient information in consensus-based optimization: On global convergence in mean-field law.” In: *arXiv preprint arXiv:2211.12184* (2022).

[Pin+17] R. Pinnau, C. Totzeck, O. Tse, and S. Martin. “A consensus-based model for global optimization and its mean-field limit.” In: *Mathematical Models and Methods in Applied Sciences* 27.01 (2017), pp. 183–204.

[Car+21] J. A. Carrillo, S. Jin, L. Li, and Y. Zhu. "A consensus-based global optimization method for high dimensional machine learning problems." In: *ESAIM: Control, Optimisation and Calculus of Variations* 27 (2021), S5.

[Cau+47] A. Cauchy et al. "Méthode générale pour la résolution des systemes d'équations simultanées." In: *Comp. Rend. Sci. Paris* 25.1847 (1847), pp. 536–538.

[RM51] H. Robbins and S. Monro. "A stochastic approximation method." In: *The annals of mathematical statistics* (1951), pp. 400–407.

[Sha+18] A. Shafahi, W. R. Huang, C. Studer, S. Feizi, and T. Goldstein. "Are adversarial examples inevitable?" In: *arXiv preprint arXiv:1809.02104* (2018).

[FFF18] A. Fawzi, H. Fawzi, and O. Fawzi. "Adversarial vulnerability for any classifier." In: *Advances in neural information processing systems* 31 (2018).

[Sta+21] J. Stanczuk, C. Etmann, L. M. Kreusser, and C.-B. Schönlieb. "Wasserstein GANs work because they fail (to approximate the Wasserstein distance)." In: *arXiv preprint arXiv:2103.01678* (2021).

[Bun+23] L. Bungert, N. G. Trillos, M. Jacobs, D. McKenzie, Đ. Nikolić, and Q. Wang. "It begins with a boundary: A geometric view on probabilistically robust learning." In: *arXiv preprint arXiv:2305.18779* (2023).

[LC10] Y. LeCun and C. Cortes. "MNIST handwritten digit database." In: (2010).

[Eng+18] L. Engstrom, B. Tran, D. Tsipras, L. Schmidt, and A. Madry. "A rotation and a translation suffice: Fooling cnns with simple transformations." In: (2018).

[Guo+17] C. Guo, M. Rana, M. Cisse, and L. Van Der Maaten. "Countering adversarial images using input transformations." In: *arXiv preprint arXiv:1711.00117* (2017).

[ANR74] N. Ahmed, T. Natarajan, and K. R. Rao. "Discrete cosine transform." In: *IEEE transactions on Computers* 100.1 (1974), pp. 90–93.

[Yua+19] X. Yuan, P. He, Q. Zhu, and X. Li. "Adversarial examples: Attacks and defenses for deep learning." In: *IEEE transactions on neural networks and learning systems* 30.9 (2019), pp. 2805–2824.

[KGB16] A. Kurakin, I. Goodfellow, and S. Bengio. "Adversarial machine learning at scale." In: *arXiv preprint arXiv:1611.01236* (2016).

[Mad+17] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu. "Towards deep learning models resistant to adversarial attacks." In: *arXiv preprint arXiv:1706.06083* (2017).

[BGM23] L. Bungert, N. García Trillos, and R. Murray. "The geometry of adversarial training in binary classification." In: *Information and Inference: A Journal of the IMA* 12.2 (2023), pp. 921–968.

[LeC+95] Y. LeCun et al. "Learning algorithms for classification: A comparison on handwritten digit recognition." In: *Neural networks: the statistical mechanics perspective* 261.276 (1995), p. 2.

[Has+20] M. Hasannasab, J. Hertrich, S. Neumayer, G. Plonka, S. Setzer, and G. Steidl. “Parseval proximal neural networks.” In: *Journal of Fourier Analysis and Applications* 26 (2020), pp. 1–31.

[Gou+20] H. Gouk, E. Frank, B. Pfahringer, and M. J. Cree. “Regularisation of neural networks by enforcing Lipschitz continuity.” In: *Machine Learning* (2020), pp. 1–24.

[KMP20] V. Krishnan, A. A. A. Makdah, and F. Pasqualetti. “Lipschitz Bounds and Provably Robust Training by Laplacian Smoothing.” In: *arXiv preprint arXiv:2006.03712* (2020).

[Sha+19] A. Shafahi, M. Najibi, M. A. Ghiasi, Z. Xu, J. Dickerson, C. Studer, L. S. Davis, G. Taylor, and T. Goldstein. “Adversarial training for free!” In: *Advances in Neural Information Processing Systems* 32 (2019).

[XRV17] H. Xiao, K. Rasul, and R. Vollgraf. “Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms.” In: *CoRR* abs/1708.07747 (2017).

[Gho+21] A. Gholami, S. Kim, Z. Dong, Z. Yao, M. W. Mahoney, and K. Keutzer. “A survey of quantization methods for efficient neural network inference.” In: *arXiv preprint arXiv:2103.13630* (2021).

[EMH19] T. Elsken, J. H. Metzen, and F. Hutter. “Neural architecture search: A survey.” In: *The Journal of Machine Learning Research* 20.1 (2019), pp. 1997–2017.

[How+17] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. “Mobilenets: Efficient convolutional neural networks for mobile vision applications.” In: *arXiv preprint arXiv:1704.04861* (2017).

[Ban+18] R. Banner, I. Hubara, E. Hoffer, and D. Soudry. “Scalable methods for 8-bit training of neural networks.” In: *Advances in neural information processing systems* 31 (2018).

[CBD14] M. Courbariaux, Y. Bengio, and J.-P. David. “Training deep neural networks with low precision multiplications.” In: *arXiv preprint arXiv:1412.7024* (2014).

[Sch92] J. Schmidhuber. “Learning complex, extended sequences using the principle of history compression.” In: *Neural Computation* 4.2 (1992), pp. 234–242.

[HVD15] G. Hinton, O. Vinyals, and J. Dean. “Distilling the knowledge in a neural network.” In: *arXiv preprint arXiv:1503.02531* (2015).

[LDS89] Y. LeCun, J. Denker, and S. Solla. “Optimal brain damage.” In: *Advances in neural information processing systems* 2 (1989).

[CFP97] G. Castellano, A. M. Fanelli, and M. Pelillo. “An iterative pruning algorithm for feedforward neural networks.” In: *IEEE transactions on Neural networks* 8.3 (1997), pp. 519–531.

[CM73] J. F. Claerbout and F. Muir. “Robust modeling with erratic data.” In: *Geophysics* 38.5 (1973), pp. 826–844.

[Tib96] R. Tibshirani. “Regression shrinkage and selection via the lasso.” In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (1996), pp. 267–288.

[Nit14] A. Nitanda. “Stochastic proximal gradient descent with acceleration techniques.” In: *Advances in Neural Information Processing Systems* 27 (2014), pp. 1574–1582.

[RVV20] L. Rosasco, S. Villa, and B. C. Vũ. “Convergence of stochastic proximal gradient algorithm.” In: *Applied Mathematics & Optimization* 82.3 (2020), pp. 891–917.

[Moc+18] D. C. Mocanu, E. Mocanu, P. Stone, P. H. Nguyen, M. Gibescu, and A. Liotta. “Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science.” In: *Nature communications* 9.1 (2018), pp. 1–12.

[DZ19] T. Dettmers and L. Zettlemoyer. “Sparse networks from scratch: Faster training without losing performance.” In: *arXiv preprint arXiv:1907.04840* (2019).

[DYJ19] X. Dai, H. Yin, and N. K. Jha. “NeST: A neural network synthesis tool based on a grow-and-prune paradigm.” In: *IEEE Transactions on Computers* 68.10 (2019), pp. 1487–1497.

[Fu+22] Y. Fu, C. Liu, D. Li, Z. Zhong, X. Sun, J. Zeng, and Y. Yao. “Exploring structural sparsity of deep networks via inverse scale spaces.” In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

[Liu+21] S. Liu, D. C. Mocanu, A. R. R. Matavalam, Y. Pei, and M. Pechenizkiy. “Sparse evolutionary deep learning with over one million artificial neurons on commodity hardware.” In: *Neural Computing and Applications* 33.7 (2021), pp. 2589–2604.

[BB18] M. Benning and M. Burger. “Modern regularization methods for inverse problems.” In: *Acta Numerica* 27 (2018), pp. 1–111.

[PB+14] N. Parikh, S. Boyd, et al. “Proximal algorithms.” In: *Foundations and trends® in Optimization* 1.3 (2014), pp. 127–239.

[Sca+17] S. Scardapane, D. Comminiello, A. Hussain, and A. Uncini. “Group sparse regularization for deep neural networks.” In: *Neurocomputing* 241 (2017), pp. 81–89.

[CP08] P. L. Combettes and J.-C. Pesquet. “Proximal thresholding algorithm for minimization over orthonormal bases.” In: *SIAM Journal on Optimization* 18.4 (2008), pp. 1351–1376.

[DDD04] I. Daubechies, M. Defrise, and C. De Mol. "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint." In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 57.11 (2004), pp. 1413–1457.

[FNW07] M. A. Figueiredo, R. D. Nowak, and S. J. Wright. "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems." In: *IEEE Journal of selected topics in signal processing* 1.4 (2007), pp. 586–597.

[Cha04] A. Chambolle. "An algorithm for total variation minimization and applications." In: *Journal of Mathematical imaging and vision* 20 (2004), pp. 89–97.

[CP11] A. Chambolle and T. Pock. "A first-order primal-dual algorithm for convex problems with applications to imaging." In: *Journal of mathematical imaging and vision* 40 (2011), pp. 120–145.

[Bre67] L. M. Bregman. "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming." In: *USSR computational mathematics and mathematical physics* 7.3 (1967), pp. 200–217.

[De 93] E. De Giorgi. "New problems on minimizing movements." In: *Ennio de Giorgi: Selected Papers* (1993), pp. 699–713.

[ROF92] L. I. Rudin, S. Osher, and E. Fatemi. "Nonlinear total variation based noise removal algorithms." In: *Physica D: nonlinear phenomena* 60.1-4 (1992), pp. 259–268.

[Bur+06] M. Burger, G. Gilboa, S. Osher, J. Xu, et al. "Nonlinear inverse scale space methods." In: *Communications in Mathematical Sciences* 4.1 (2006), pp. 179–212.

[Bur+07] M. Burger, K. Frick, S. Osher, and O. Scherzer. "Inverse total variation flow." In: *Multiscale Modeling & Simulation* 6.2 (2007), pp. 366–395.

[Yin+08] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. "Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing." In: *SIAM Journal on Imaging sciences* 1.1 (2008), pp. 143–168.

[COS09] J.-F. Cai, S. Osher, and Z. Shen. "Convergence of the linearized Bregman iteration for ℓ_1-norm minimization." In: *Mathematics of Computation* 78.268 (2009), pp. 2127–2136.

[Vil+23] S. Villa, S. Matet, B. C. Vũ, and L. Rosasco. "Implicit regularization with strongly convex bias: Stability and acceleration." In: *Analysis and Applications* 21.01 (2023), pp. 165–191.

[BT03] A. Beck and M. Teboulle. "Mirror descent and nonlinear projected subgradient methods for convex optimization." In: *Operations Research Letters* 31.3 (2003), pp. 167–175.

[NY83] A. S. Nemirovskij and D. B. Yudin. "Problem complexity and method efficiency in optimization." In: (1983).

[Nem+09] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. "Robust stochastic approximation approach to stochastic programming." In: *SIAM Journal on optimization* 19.4 (2009), pp. 1574–1609.

[Nes83] Y. Nesterov. "A method for unconstrained convex minimization problem with the rate of convergence o (1/k2)." In: *Doklady ANSSSR* 269.3 (1983), pp. 543–547.

[Qia99] N. Qian. "On the momentum term in gradient descent learning algorithms." In: *Neural networks* 12.1 (1999), pp. 145–151.

[KB14] D. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization." In: *arXiv preprint arXiv:1412.6980* (2014).

[HR21] F. Hanzely and P. Richtárik. "Fastest rates for stochastic mirror descent methods." In: *Computational Optimization and Applications* 79 (2021), pp. 717–766.

[ZH18] S. Zhang and N. He. "On the convergence rate of stochastic mirror descent for nonsmooth nonconvex optimization." In: *arXiv preprint arXiv:1806.04781* (2018).

[DOr+21] R. D'Orazio, N. Loizou, I. Laradji, and I. Mitliagkas. "Stochastic mirror descent: Convergence analysis and adaptive variants via the mirror stochastic Polyak stepsize." In: *arXiv preprint arXiv:2110.15412* (2021).

[AKL22] P.-C. Aubin-Frankowski, A. Korba, and F. Léger. "Mirror descent with relative smoothness in measure spaces, with application to sinkhorn and em." In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 17263–17275.

[Ben+21] M. Benning, M. M. Betcke, M. J. Ehrhardt, and C.-B. Schönlieb. "Choose your path wisely: gradient descent in a Bregman distance framework." In: *SIAM Journal on Imaging Sciences* 14.2 (2021), pp. 814–843.

[HZ93] G. E. Hinton and R. Zemel. "Autoencoders, minimum description length and Helmholtz free energy." In: *Advances in neural information processing systems* 6 (1993).

[KLM21] N. Kovachki, S. Lanthaler, and S. Mishra. "On universal approximation and error bounds for Fourier neural operators." In: *The Journal of Machine Learning Research* 22.1 (2021), pp. 13237–13312.

[HW62] D. H. Hubel and T. N. Wiesel. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." In: *The Journal of physiology* 160.1 (1962), p. 106.

[Kov+21] N. B. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. M. Stuart, and A. Anandkumar. "Neural Operator: Learning Maps Between Function Spaces." In: *arXiv:2108.08481* (2021).

[Bri19] T. Briand. “Trigonometric Polynomial Interpolation of Images.” In: *Image Processing On Line* 9 (Oct. 2019), pp. 291–316.

[HG16] D. Hendrycks and K. Gimpel. “Gaussian Error Linear Units (GELUs).” In: *arXiv:1606.08415* (2016).

Theses

[Roi21] T. Roith. “Master thesis: Continuum limit of Lipschitz learning on graphs.” MA thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg, 2021.

[Sma10] C. K. Smart. “On the infinity Laplacian and Hrushovski’s fusion.” PhD thesis. UC Berkeley, 2010.

[Kab22] S. Kabri. “Fourier Neural Operators for Image Classification.” MA thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg, 2022.

Online

[Pio21] G. Piosenka. *BIRDS 500 - SPECIES IMAGE CLASSIFICATION*. 2021. URL: %7Bhttps://www.kaggle.com/datasets/gpiosenka/100-bird-species%7D.