

Meet my Expectations: On the Interplay of Trustworthiness and Deep Learning Optimization

Dissertation

der Fakultät für Mathematik und Informatik
der Universität des Saarlandes
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Adrián Javaloy Bornás, M. Sc.

Saarbrücken
2024

Gedruckt mit Genehmigung der Fakultät für Mathematik und Informatik der Universität des Saarlandes.

Tag des Kolloquiums: 11. März 2025

Dekan: Prof. Dr. Roland Speicher

Prüfungsausschuss:

Vorsitz:	Prof. Dr. Martina Maggio
1. Berichterstatter:	Prof. Dr. Isabel Valera
2. Berichterstatter:	Prof. Dr. Jilles Vreeken
3. Berichterstatter:	Prof. Dr. Benjamin Bloem-Reddy
Akademischer Beisitzer:	Dr. Kavya Gupta

Dedicated to Alfonso Bornás Bayonas.

Your family will never forget you.

Disclaimer. This style is a modified version of Felix Dangel's thesis template, which in turn uses Federico Marotta's kaobook template based on Ken Arroyo Ohori's doctoral thesis, and I am grateful to all of them for their openness to share their \LaTeX codes. Otherwise, this thesis would not look nearly the way it does.

Acknowledgments

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum. Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum. Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

Thank you!

Felix Dangel
Tübingen, August 31, 2022

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Zusammenfassung

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muß keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Das hier ist der zweite Absatz. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muß keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Und nun folgt – ob man es glaubt oder nicht – der dritte Absatz. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muß keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Nach diesem vierten Absatz beginnen wir eine neue Zählung. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muß keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muß keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Publications

Plagiarism is the sincerest form of flattery.

Oscar Wilde

The following publications are at the core of my PhD research and covered in this dissertation:

- [I] ‘Lipschitz standardization for multivariate learning.’
Javaloy, A., and Valera, I., in arXiv, 2020.
- [II] ‘RotoGrad: Gradient Homogenization in Multitask Learning.’
Javaloy, A., and Valera, I., in ICLR (👁), 2022.
- [III] ‘Mitigating Modality Collapse in Multimodal VAEs via Impartial Optimization.’
Javaloy, A., Meghdadi, M., and Valera, I., in ICML (👁), 2022.
- [IV] ‘Causal normalizing flows: from theory to practice.’
Javaloy, A., Sánchez-Martín, P., and Valera, I., in NeurIPS (🗣), 2023.

The following publications originated during my time as a PhD student but are omitted from this thesis:

- [V] ‘Text normalization using encoder–decoder networks based on the causal feature extractor.’
Javaloy, A., and García-Mateos, G., in Applied Sciences, 2020.
- [VI] ‘Preliminary results on different text processing tasks using encoder-decoder networks and the causal feature extractor.’
Javaloy, A., and García-Mateos, G., in Applied Sciences, 2020.
- [VII] ‘Relative gradient optimization of the Jacobian term in unsupervised deep learning.’
Gresele, L., Fissore, G., **Javaloy, A.**, Schölkopf, B., and Hyvärinen, A., in NeurIPS, 2020.
- [VIII] ‘Learnable Graph Convolutional Attention Networks.’
Javaloy, A., Sánchez-Martín, P., Levi, A., and Valera, I., in ICLR, 2023.

*: Equal contribution; 🗣: Oral; 👁: Spotlight.

Table of Contents

Acknowledgments	v
Abstract	vii
Zusammenfassung	ix
Publications	xi
Table of Contents	xiii
Notation	xv
1. Overview	1
1.1. Introduction	1
1.2. Outline	3
I. BACKGROUND & MOTIVATION	5
2. Background 1	7
2.1. Background topic 1	7
2.2. Background topic 2	7
3. Background 2 (advanced)	9
3.1. Background topic 1 (advanced)	9
3.2. Background topic 2 (advanced)	9
II. BACKPROPAGATION BEYOND THE GRADIENT	11
4. Paper 1	13
4.1. Introduction	13
4.2. Theory & Implementation	13
4.3. Evaluation & Benchmarks	13
4.4. Experiments	13
4.5. Conclusion	13
5. Paper 2	15
5.1. Introduction & Motivation	16
5.2. Cockpit's Instruments	16
5.3. Experiments	16
5.4. Showcase	16
5.5. Benchmark	16
5.6. Conclusion	16
III. CONCLUSION & FUTURE DIRECTIONS	17
6. Conclusion & Future Directions	19
6.1. Summary & Impact	20
6.2. Future Work	21

IV. APPENDIX	23
A. Additional Material for Chapter 4	25
B. Additional Material for Chapter 5	27
Bibliography	29

Notation

The notation is influenced by Goodfellow et al. [2].

Tensors, Matrices, Vectors, Numbers

a	A scalar
\mathbf{a}	A column vector
A	A matrix
\mathbf{A}	A tensor
a_i or $[\mathbf{a}]_i$	The i th entry of the vector \mathbf{a}
$A_{i,j}$ or $[A]_{i,j}$	The (i, j) th entry of the matrix A (row i , column j)
$[A]_{i,:}$ (or $[A]_{:,j}$)	The i th row (or j th column) of the matrix A
$A_{i,j,k}$ or $[\mathbf{A}]_{i,j,k}$	The (i, j, k) th entry of the tensor \mathbf{A}
$\text{vec}(A), \text{vec}(\mathbf{A})$	Matrix/tensor flattened into a vector; convention implies $\text{vec}(ABC) = (C^\top \otimes A) \text{vec}(B)$
$\text{diag}(\mathbf{a})$	The square matrix with vector \mathbf{a} on the diagonal and zeros elsewhere
$\text{diag}(A)$	The vector containing the diagonal elements of the matrix A
$\text{diag}(A_1, \dots, A_L)$	A block-diagonal matrix with diagonal blocks given by square matrices A_1, \dots, A_L
$\text{Tr}(A), \det(A)$	Trace and determinant of a matrix A
$\ \mathbf{a}\ _2$	L_2 norm of vector \mathbf{a} , i.e. $\ \mathbf{a}\ _2^2 = \mathbf{a}^\top \mathbf{a}$
$\text{eig}(A) := \{(\lambda_k, \mathbf{e}_k)\}_k$	Eigendecomposition of the matrix A , eigenpairs $(\lambda_k, \mathbf{e}_k)$ satisfy $A\mathbf{e}_k = \lambda_k \mathbf{e}_k$
$(\lambda_k(A), \mathbf{e}_k(A))$	k th eigenpair (eigenvalue, eigenvector) of matrix A
$A \otimes B$	Kronecker product of two matrices, For two vectors \mathbf{a}, \mathbf{b} , one has $\mathbf{a} \otimes \mathbf{b}^\top = \mathbf{a}\mathbf{b}^\top$
$\mathbf{A} \odot \mathbf{B}, A \odot B, \mathbf{a} \odot \mathbf{b}$	Elementwise multiplication (Hadamard product) of two tensors, matrices, vectors
$\mathbf{A} \oslash \mathbf{B}, A \oslash B, \mathbf{a} \oslash \mathbf{b}$	Elementwise division (Hadamard division) of two tensors, matrices, vectors
$\mathbf{A}^{\odot 2}, A^{\odot 2}, \mathbf{a}^{\odot 2}$	Elementwise square of a tensor, matrix, vector
$\mathbf{A}^{\odot 1/2}, A^{\odot 1/2}, \mathbf{a}^{\odot 1/2}$	Elementwise square root of a tensor, matrix, vector

Empirical Risk Minimization

A datum is usually indicated by a subscript n .

(\mathbf{x}, \mathbf{y})	Labeled datum with input features \mathbf{x} and target \mathbf{y}
$(\mathbf{x}_n, \mathbf{y}_n)$	Datum n from a dataset
D	Total number of parameters in a model
$\boldsymbol{\theta} \in \Theta := \mathbb{R}^D$	Parameter vector of a model
$\mathbf{f} := f_{\boldsymbol{\theta}}(\mathbf{x})$	Prediction of a model $f_{\boldsymbol{\theta}}$ for input features \mathbf{x}
ℓ or $\ell(\mathbf{f}, \mathbf{y})$	Loss function to compare prediction and target; convex in \mathbf{f}
$\mathbb{D} := \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^{ \mathbb{D} }$	A dataset containing instances of labeled data $(\mathbf{x}_n, \mathbf{y}_n)$ indexed by n
\mathbb{B}	A mini-batch $\mathbb{B} \subseteq \mathbb{D}$
N	Number of data in a mini-batch or a dataset, depending on the context
$\mathbf{f}_n := f_{\boldsymbol{\theta}}(\mathbf{x}_n)$	Model prediction for datum n
ℓ_n or $\ell(\mathbf{f}_n, \mathbf{y}_n)$	Loss of datum n
$p_{\mathbb{D}}(\mathbf{x}, \mathbf{y})$	Empirical distribution of a dataset \mathbb{D}
$\mathcal{L}_{\mathbb{D}}(\boldsymbol{\theta})$	Empirical risk implied by the empirical distribution of a dataset \mathbb{D}
$\mathcal{L}_{\mathbb{D}_{\text{train}}}(\boldsymbol{\theta}), \mathcal{L}_{\mathbb{B}}(\boldsymbol{\theta}), \text{etc.}$	Training loss, mini-batch loss, etc.

Neural Networks

The layer number is indicated by parenthesized superscripts $^{(l)}$.

L	Total number of layers
$d^{(l)}$	Number of parameters in layer l ; total number of parameters is $D = \sum_{l=1}^L d^{(l)}$
$\theta^{(l)} \in \mathbb{R}^{d^{(l)}}$	Parameter vector of layer l , potentially empty for parameter-free layers like activations
$h^{(l-1)}$	Number of (hidden) inputs fed into layer l
$M := h^{(0)}, C := h^{(L)}$	Input feature dimension, output dimension (number of classes for classification)
$\mathbf{z}^{(l-1)} \in \mathbb{R}^{h^{(l-1)}}$	(Hidden) features fed into layer l (output of layer $l-1$)
$\mathbf{x} := \mathbf{z}^{(0)}, \mathbf{f} := \mathbf{z}^{(L)}$	Input to the neural network, and its prediction for input \mathbf{x}
$f_{\theta^{(l)}}^{(l)}$	Layer l parameterized by $\theta^{(l)}$, mapping input $\mathbf{z}^{(l-1)}$ to output $\mathbf{z}^{(l)}$
$f_{\theta} := f_{\theta^{(L)}} \circ \dots \circ f_{\theta^{(1)}}^{(1)}$	Sequential feedforward neural network parameterized by θ , maps input \mathbf{x} to output \mathbf{f}
θ	Parameter vector, concatenation of parameters over layers $\theta := (\theta^{(1)\top}, \dots, \theta^{(L)\top})^\top$

Derivatives

∇, \mathbf{J} and ∇^2 denote the gradient, Jacobian, and Hessian, respectively.

$\mathbf{J}_a \mathbf{b}$	Jacobian matrix of a vector \mathbf{b} w.r.t. a vector \mathbf{a} , $[\mathbf{J}_a \mathbf{b}]_{i,j} = \partial[\mathbf{b}]_i / \partial[\mathbf{a}]_j$
$\mathbf{J}_A \mathbf{B}$	Generalized Jacobian matrix for tensor variables, $[\mathbf{J}_A \mathbf{B}]_{i,j} = \partial[\text{vec} \mathbf{B}]_i / \partial[\text{vec} \mathbf{A}]_j$
$\nabla_a b := (\mathbf{J}_a b)^\top$	Gradient vector of a scalar b w.r.t. a vector \mathbf{a} , $[\nabla_a b]_i = \partial b / \partial a_i$
$\nabla_a^2 b$	Hessian matrix of a scalar b w.r.t. a vector \mathbf{a} , $[\nabla_a^2 b]_{i,j} = \partial^2 b / \partial[\mathbf{a}]_i \partial[\mathbf{a}]_j$ (symmetric)
$\nabla_a^2 \mathbf{b}, \nabla_A^2 \mathbf{B}$	Generalized Hessian matrix (in general not quadratic, hence not symmetric) of a vector \mathbf{b} w.r.t. a vector \mathbf{a} , or more general tensor variables
$\mathbf{g}_n(\theta) := \nabla_{\theta} \ell_n(\theta)$	Gradient of the loss implied by sample n
$\mathbf{g}_{\mathbb{D}}(\theta) := \nabla_{\theta} \mathcal{L}_{\mathbb{D}}(\theta)$	Gradient of the empirical risk implied by a dataset \mathbb{D}
$\mathbf{g}_{\mathbb{B}}(\theta) := \nabla_{\theta} \mathcal{L}_{\mathbb{B}}(\theta)$	Mini-batch gradient
$\mathbf{H}_n(\theta) := \nabla_{\theta}^2 \ell_n(\theta)$	Hessian of the loss implied by sample n
$\mathbf{H}_{\mathbb{D}}(\theta) := \nabla_{\theta}^2 \mathcal{L}_{\mathbb{D}}(\theta)$	Hessian of the empirical risk implied by a dataset \mathbb{D}
$\mathbf{H}_{\mathbb{B}}(\theta) := \nabla_{\theta}^2 \mathcal{L}_{\mathbb{B}}(\theta)$	Mini-batch Hessian
$\mathbf{H}^{(l)}(\theta^{(l)})$ or $\mathbf{H}(\theta^{(l)})$	The block in the Hessian corresponding to layer l
$\mathbf{G}_{\mathbb{D}}(\theta)$	Generalized Gauss-Newton matrix on a dataset \mathbb{D}
$\mathbf{G}^{(l)}(\theta^{(l)})$ or $\mathbf{G}(\theta^{(l)})$	The block in the generalized Gauss-Newton matrix corresponding to layer l

Statistics

$\mathcal{U}(\{1, \dots, N\})$	Uniform distribution over $\{1, \dots, N\}$
$\mathcal{N}(x \mid \mu, \sigma^2)$	Uni-variate normal/Gaussian distribution of random variable x , with mean μ , positive variance σ^2 , and density $\mathcal{N}(x \mid \mu, \sigma^2) = 1/\sigma\sqrt{2\pi} \exp[-1/2((x-\mu)/\sigma)^2]$
$\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$	Multi-variate normal/Gaussian distribution of random vector \mathbf{x} with mean vector $\boldsymbol{\mu}$, PSD covariance matrix $\boldsymbol{\Sigma}$, and density $\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = 1/(\sqrt{2\pi} \det \boldsymbol{\Sigma}) \exp[-1/2(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})]$
$\text{Cat}(c \mid \mathbf{p})$	Multinomial/Categorical distribution with probabilities \mathbf{p} for categories c

Miscellaneous

\log	The natural logarithm (base e, i.e. $\log(e) = 1$)
$\text{onehot}(c)$	One-hot vector of class c with $\text{onehot}(c) = \delta_{i,c}$
$\text{softmax}(\mathbf{a})$	Softmax probabilities of the logits \mathbf{a} , $[\text{softmax}(\mathbf{a})]_c = \exp(a_c) / \sum_{i=1} \exp(a_i)$.
$\delta_{i,j}, \delta(\mathbf{x} - \mathbf{a})$	Kronecker delta ($\delta_{i,i} = 1$ and $\delta_{i,j \neq i} = 0$), Dirac delta distribution
$(\mathbb{X} \rightarrow \mathbb{Y})$	Signature of a function that maps between \mathbb{X} and \mathbb{Y}
$\{\mathbf{x}_n\}$ or $\{\mathbf{x}_n\}_n$	A set/collection of vectors $\mathbf{x}_1, \mathbf{x}_2, \dots$ over the index set implied by n
$\hat{\mathbf{e}}_i$	Unit vector in direction i , i.e. $\hat{\mathbf{e}}_i = \text{onehot}(i)$

$\mathbf{1}_m$	An m -dimensional vector containing ones everywhere
$\log(\mathbf{a}), \exp(\mathbf{a})$	Elementwise natural logarithm and exponential function of a vector
$m_{\theta_t}(\theta)$	Local approximation of the loss in around θ_t

Acronyms & Abbreviations

<i>E.g.</i> or <i>e.g.</i>	For example (<i>exempli gratia</i>)
<i>Etc.</i> or <i>etc.</i>	And so on (<i>et cetera</i>)
<i>I.e.</i> or <i>i.e.</i>	That is (<i>id est</i>)
<i>I.i.d.</i> or <i>i.i.d.</i>	Independent and identically distributed
<i>W.r.t.</i> or <i>w.r.t.</i>	With respect to
AD	Automatic differentiation
API	Application Programming Interface
BDA	Block diagonal approximation
CG	Conjugate gradients
CNN	Convolutional neural network
CPU	Central processing unit
DNN	Deep neural network
DP	Differential privacy
FCNN	Fully-connected neural network
GGN	Generalized Gauss-Newton (matrix)
GN	Gauss-Newton (matrix)
GPU	Graphics processing unit
HBP	Hessian backpropagation
JMP	Jacobian-matrix product
JVP	Jacobian-vector product
KFAC	Kronecker-factored curvature
KFC	Kronecker factors for convolution
KFLR	Kronecker-factored low rank
KFRA	Kronecker-factored recursive approximation
MAP	Maximum a posteriori (estimation)
MC	Monte Carlo
MJP	Matrix-Jacobian product
ML	Machine learning
MLE	Maximum likelihood estimation
MLP	Multi-layer perceptron
NGD	Natural gradient descent
PCH	Positive-curvature Hessian
PD	Positive definite
PSD	Positive semi-definite
ResNet	Residual (neural) network
SNR	Signal-to-noise ratio
TPU	Tensor processing unit
VJP	Vector-Jacobian product

Venceréis, pero no convenceréis.

Miguel de Unamuno, 1936

1.1 Introduction	1
1.2 Outline	3

1.1 Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

1.1.1 Subsection 1	2
1.1.2 Subsection 2	2

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet

aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Algorithm 1.1: Canonical deep learning training loop. After setting up the data, model, loss function, and optimizer, iterate over batches: in each iteration, compute the mini-batch loss in a forward pass, and its gradient with a backward pass. Then use the gradient as learning signal to update the model parameters.

```

1 % dataset = ...      # Learning task examples
2
3 model = ...          # Practitioner's choice
4 loss_func = ...      # Practitioner's choice
5
6 optimizer = ...      # First-order method
7
8 while not_converged: # Standard training loop
9     features, targets = dataset.next_minibatch()
10
11     # Forward pass: Compute the loss
12     predictions = model(features)
13     loss = loss_func(predictions, targets)
14
15     # Backward pass: Compute the gradient
16     loss.backward()
17
18     # Update model parameters using the gradient
19     optimizer.step()
20     optimizer.zero_grad()

```

Reference to [Section 1.1](#).

1.1.1 Subsection 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

1.1.2 Subsection 2

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque

placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

1.2 Outline

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo,

lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

You have to run one the script `build-force.sh` to generate the following externalized TikZ figure:

Figure 1.1: A TikZ figure: This figure checks if externalization works.

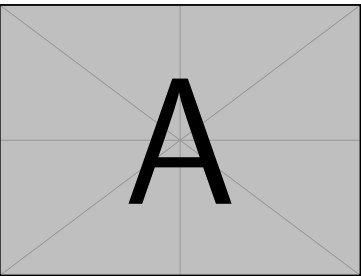
[[IMAGE DISCARDED DUE TO '/tikz/external/mode=list and make']]

Disclaimer 1.1
Chapter 4 is based on the peer-reviewed conference publication with the following co-author contributions:

Dangel, F., Kunstner, F., and Hennig, P., ‘BackPACK: Packing more into Backprop.’ *International Conference on Learning Representations (ICLR)*. 2020 [1]

	Ideas	Experiments	Analysis	Writing
F. Dangel	33 %	55 %	45 %	35 %
F. Kunstner	33 %	45 %	45 %	45 %
P. Hennig	33 %	0 %	10 %	20 %

Chapter 4: BackPACK: an efficient framework built on top of PyTorch that extends the backpropagation algorithm.



github.com/f-dangel/backpack

Part I.

Background & Motivation

Background 1

2.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.1 Background topic 1

Definition 2.1 (Hessian) Let $b : \mathbb{R}^D \rightarrow \mathbb{R}; a \mapsto b(a)$ be a differentiable vector-to-scalar function. The Hessian $\nabla_a^2 b \in \mathbb{R}^{D \times D}$ of b w.r.t. a is a symmetric matrix containing the second-order partial derivatives

$$\nabla_a^2 b = \frac{\partial^2 b}{\partial a \partial a^\top} \quad \text{with} \quad [\nabla_a^2 b]_{i,j} = \frac{\partial^2 b}{\partial a_i \partial a_j} \quad (2.2)$$

The Hessian will often be denoted by H . E.g. $H_{p_{\text{data}}}(\theta) := \nabla_\theta^2 \mathcal{L}_{p_{\text{data}}}(\theta)$ for the Hessian of the population risk, and $H_{\mathbb{D}}(\theta) := \nabla_\theta^2 \mathcal{L}_{\mathbb{D}}(\theta)$ for the Hessian of the empirical risk on a dataset \mathbb{D} (with $\mathbb{D} = \mathbb{D}_{\text{train}}, \mathbb{B}$ for the train loss and mini-batch Hessian).

2.1 Background topic 1 7

2.2 Background topic 2 7

The arrangement of partial derivatives in the generalizations of Jacobian and Hessian implies the following chain rule generalization for second-order derivatives:

Theorem 2.1 (Chain rule for the generalized Hessian) Let $b : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $c : \mathbb{R}^m \rightarrow \mathbb{R}^p$ be twice differentiable and $d = c \circ b : \mathbb{R}^n \rightarrow \mathbb{R}^p, a \mapsto d(a) = c(b(a))$. The relation between the Hessian of d and the Jacobians and Hessians of the constituents c and b is given by

$$\begin{aligned} \nabla_a^2 d(a) &= [I_p \otimes J_a b(a)]^\top [\nabla_b^2 c(b)] J_a b(a) \\ &\quad + [J_b c(b) \otimes I_n] \nabla_a^2 b(a) \end{aligned} \quad (2.1)$$

2.2 Background topic 2

Example 2.1 (Least squares regression & square loss) Regression associates features in $\mathbb{X} = \mathbb{R}^M$ with targets in $\mathbb{Y} = \mathbb{R}^C$. A prediction in $\mathbb{F} = \mathbb{R}^C$ compares to its ground truth via the mean squared error¹

$$\ell(f, y) = \frac{1}{C} \sum_{c=1}^C (y_c - f_c)^2 = \frac{1}{C} \|y - f\|_2^2 \quad (2.3)$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

1: There exist different conventions for the normalization factor. This text adapts the implementation of `MSELoss` (with `reduction="mean"` mode) in PyTorch for consistency with the code presented in later chapters. Normalizing by $1/C$ is also close to what the name, mean squared error, suggests.

Remark 2.1 (The log-probability's θ -gradient vanishes in expectation)

$$\begin{aligned} & - \int_{\Omega} p_{\theta}(z) \nabla_{\theta} \log p_{\theta}(z) dz \\ &= - \int_{\Omega} p_{\theta}(z) \frac{\nabla_{\theta} p_{\theta}(z)}{p_{\theta}(z)} dz \\ &= - \nabla_{\theta} \left(\int_{\Omega} p_{\theta}(z) dz \right) \\ &= - \nabla_{\theta} 1 = 0 \end{aligned}$$

Table 2.1: Forward pass for common modules used in feedforward networks.

Input and output are denoted x, z rather than $z^{(l)}, z^{(l+1)}$ to avoid clutter. I is the identity matrix. Bold upper-case symbols (W, X, Z, \dots) denote matrices and bold upper-case sans serif symbols ($\mathbf{W}, \mathbf{X}, \mathbf{Z}, \dots$) denote tensors.

OPERATION	FORWARD
Matrix-vector multiplication	$z(x, W) = Wx$
Matrix-matrix multiplication	$Z(X, W) = WX$
Addition	$z(x, b) = x + b$
Elementwise activation	$z(x) = \phi(x)$, s.t. $z_i(x) = \phi(x_i)$
Skip-connection	$z(x, \theta) = x + s(x, \theta)$
Reshape/view	$Z(X) = \text{reshape}(X)$
Index select/map π	$z(x) = \Pi x, \Pi_{j, \pi(j)} = 1,$
Convolution	$Z(X, W) = X \star W,$ $Z(W, \llbracket X \rrbracket) = W \llbracket X \rrbracket,$
Square loss	$\ell(f, y) = 1/C (y - f)^\top (y - f)$
Softmax cross-entropy	$\ell(f, y) = -\text{onehot}(y)^\top \log [p(f)]$

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Background 2 (advanced)

3.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

- 3.1 Background topic 1 (advanced) 9
- 3.2 Background topic 2 (advanced) 9

3.1 Background topic 1 (advanced)

3.2 Background topic 2 (advanced)

Part II.

Backpropagation Beyond the Gradient

Paper 1

4.

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

4.1 Introduction

4.2 Theory & Implementation

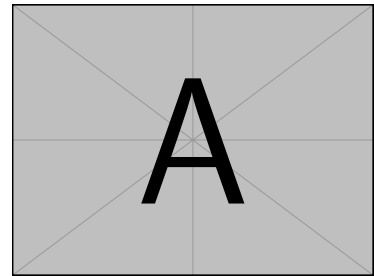
4.3 Evaluation & Benchmarks

4.4 Experiments

4.5 Conclusion

4.1 Introduction	13
4.2 Theory & Implementation .	13
4.3 Evaluation & Benchmarks .	13
4.4 Experiments	13
4.5 Conclusion	13

Code and experiments available at the
Github repositories
[f-dangel/backpack](#),
[f-dangel/backpack-experiments](#)



Paper 2

5.

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

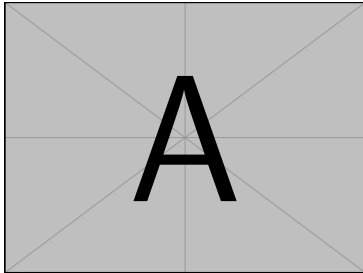
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus

5.1 Introduction & Motivation	16
5.2 Cockpit's Instruments . . .	16
5.3 Experiments	16
5.4 Showcase	16
5.5 Benchmark	16
5.6 Conclusion	16

Code and experiments available at the Github repositories [f-dangel/cockpit](#), [f-dangel/cockpit-experiments](#)



nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

5.1 Introduction & Motivation

5.2 Cockpit's Instruments

5.3 Experiments

5.4 Showcase

5.5 Benchmark

5.6 Conclusion

Acknowledgments

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Part III.

Conclusion & Future Directions

Conclusion & Future Directions

6.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis

6.1 Summary & Impact 20
6.2 Future Work 21

fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

6.1 Summary & Impact

Extending Backpropagation to the Hessian

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Packing More into Backprop

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Enabling a Closer Look Into Neural Nets

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Enabling Novel Ways to Compute with Curvature

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent

imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

6.2 Future Work

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Extending Cockpit

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Noise-aware Second-order Methods

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Optimizing Run Time & Advancing Automatic Differentiation

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Part IV.

Appendix

Additional Material for Chapter 4

A.

Additional Material for Chapter 5

B.

Bibliography

- [1] Dangel, F., Kunstner, F., and Hennig, P., 'BackPACK: Packing more into Backprop.' *International Conference on Learning Representations (ICLR)*. 2020.
- [2] Goodfellow, I. J., Bengio, Y., and Courville, A., 'Deep Learning.' 2016.