CARPATHIAN J. MATH.
Volume **42** (2026), No. 3,
Pages 541-554

Online version at `https://www.carpathian.cunbm.utcluj.ro/`
Print Edition: ISSN 1584 - 2851; Online Edition: ISSN 1843 - 4401
DOI: https://doi.org/10.37193/CJM.2026.03.04

Duality of Natural and Euclidean Gradients on the Multinomial Statistical Manifold

IOANA RĂDULESCU (LĂZĂRESCU)

ABSTRACT. This paper explores the duality of the Natural Gradient and Euclidean Gradient in the statistical manifold of multinomial distributions by examining the relationship between these gradients in dual coordinate spaces. We derive the canonical exponential form of the multinomial distribution and compute the Fisher metric using a change of basis method. The duality between the Natural and Euclidean Gradients in these spaces is demonstrated through both computational derivations and experimental validation. In a small-scale experiment, we compare the convergence rates of these gradients, confirming their duality and highlighting the practical advantages of the Natural Gradient in optimization using gradient descent methods.

ACKNOWLEDGMENTS

The author would like to thank Professor Adela Mihai for the insightful discussions and advice given during the preparation of this manuscript.

REFERENCES

[1] Amari, S.; Park, H.; Fukumizu, K. Adaptive method of realizing natural gradient learning for multilayer perceptrons. *Neural Comput.* **12** (2000), no. 6, 1399–1409.
[2] Amari, S. *Information Geometry and Its Applications*. Applied Mathematical Sciences, 194. Springer, Tokyo, 2016.
[3] Amari, S. Natural gradient works efficiently in learning. *Neural Comput.* **10** (1998), no. 2, 251–276.
[4] Amari, S. Neural learning in structured parameter spaces – natural Riemannian gradient. *Adv. Neural Inf. Process. Syst.* **9** (1996), 127–133.
[5] Amari, S.; Nagaoka, H. *Methods of Information Geometry*. Translated from the 1993 Japanese original by Daishi Harada. Translations of Mathematical Monographs, 191. American Mathematical Society, Providence, RI; Oxford University Press, Oxford, 2000.
[6] Bishop, C. M. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, 2006.
[7] Branea, I. A.; Rădulescu (Lăzărescu), I. A geometric approach of probability distributions. *Rom. J. Math. Comput. Sci.* **13** (2023), no. 1, 26–36.
[8] Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2016.
[9] Gudmundsson, S. An introduction to Riemannian geometry. Lecture notes (version 1.235, 9 December 2004), Lund University, 2004. `http://www.matematik.lu.se/matematiklu/personal/sigma/Riemann.pdf`.
[10] Harris, C. R.; Millman, K. J.; van der Walt, S. J. et al. Array programming with NumPy. *Nature* **585** (2020), no. 7825, 357–362.
[11] Loaiza O., G. I.; Cadavid M., C. A.; Arango P., J. C. Base change for Fisher metrics: case of the q-Gaussian inverse distribution. *International Journal of Mathematical and Computational Sciences* **16** (2022), no. 9, 74–80. Available at `https://publications.waset.org/10012676.pdf`.
[12] Martens, J. New insights and perspectives on the natural gradient method. *J. Mach. Learn. Res.* **21** (2020), no. 146, 1–76.
[13] Murphy, K. P. *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA, 2012.

Received: 15.08.2025. In revised form: 28.09.2025. Accepted: 16.10.2025
2020 *Mathematics Subject Classification.* 53B12, 65C20.
Key words and phrases. *Information Geometry, Multinomial Distribution, Fisher Metric, Natural Gradient.*

[14] Nielsen, F. A note on the natural gradient and its connections with the Riemannian gradient, the mirror descent, and the ordinary gradient. Preprint, 2020. Available at https://www.researchgate.net/publication/343826820_A_note_on_the_natural_gradient_and_its_connections_with_the_Riemannian_gradient_the_mirror_descent_and_the_ordinary_gradient.

[15] Nielsen, F. An elementary introduction to information geometry. *Entropy* **22** (2020), no. 10, 1100.

[16] Osawa, K.; Tsuji, Y.; Ueno, Y.; Naruse, A.; Yokota, R.; Matsuoka, S. Large-scale distributed second-order optimization using Kronecker-factored approximate curvature for deep convolutional neural networks. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), 12359–12367.

[17] Pascanu, R.; Bengio, Y. Revisiting natural gradient for deep networks. Preprint, 2014. Available at https://arxiv.org/abs/1301.3584.

[18] Pauloski, J. G.; Huang, Q.; Huang, L.; Venkataraman, S.; Chard, K.; Foster, I. T.; Zhang, Z. Kaisa: An adaptive second-order optimizer framework for deep neural networks. In: *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (SC21), Association for Computing Machinery, New York, 2021, Article 13, 1–14.

[19] Pedregosa, F.; Varoquaux, G.; Gramfort, A. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12** (2011), 2825–2830.

[20] Rădulescu (Lăzărescu), I.; Băicoianu, A.; Mihai, A. Conformal transformation of kernels: a geometric perspective on text classification. *Facta Univ. Ser. Math. Inform.*, to appear. Preprint, 2024. Available at https://arxiv.org/abs/2406.00499.

[21] Reyad, M.; Sarhan, A. M.; Arafa, M. A modified Adam algorithm for deep neural network optimization. *Neural Comput. Appl.* **35** (2023), no. 23, 17095–17112.

[22] Shima, H. *The Geometry of Hessian Structures*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2007.

[23] Shrestha, R. Natural gradient methods: perspectives, efficient-scalable approximations, and analysis. Preprint, 2023. Available at https://arxiv.org/abs/2303.05473.

[24] Udrişte, C.; Călin, O. *Geometric Modeling in Probability and Statistics*. Springer, Cham, 2014.

[1] Interdisciplinary Doctoral School, Transilvania University of Braşov, Bd. Eroilor 29, Braşov, 500036, Romania,

Email address: ioana.radulescu@unitbv.ro