Proceedings of Machine Learning Research vol 196:1–3, 2024 37th Annual Conference on Learning Theory

Contraction of Markovian Operators in Orlicz Spaces and Error Bounds for Markov Chain Monte Carlo (Extended Abstract)

Amedeo Roberto Esposito AMEDEOROBERTO.ESPOSITO@IST.AC.AT
Institute of Science and Technology Austria
Marco Mondelli MARCO.MONDELLI@IST.AC.AT
Institute of Science and Technology Austria

Editors: Shipra Agrawal and Aaron Roth

Abstract

We introduce a novel concept of convergence for Markovian processes within Orlicz spaces, extending beyond the conventional approach associated with L_p spaces. After showing that Markovian operators are contractive in Orlicz spaces, our technical contribution is an upper bound on their contraction coefficient, which admits a closed-form expression. The bound is tight in some settings, and it recovers well-known results, such as the connection between contraction and ergodicity, ultra-mixing and Doeblin's minorisation. Moreover, we can define a notion of convergence of Markov processes in Orlicz spaces, which depends on the corresponding contraction coefficient.

The key novelty comes from duality considerations: the convergence of a Markovian process determined by K depends on the contraction coefficient of its dual $K^\star$, which can in turn be bounded by considering appropriate nested norms of densities of $K^\star$ with respect to the stationary measure. Our approach stands out as the first of its kind, as it does not rely on the existence of a spectral gap. Specialising our approach to L_p spaces leads to a significant improvement upon classical Riesz-Thorin's interpolation methods. We present the following applications of the proposed framework:

1. Tighter bounds on the mixing time of Markovian processes: one can relate the contraction coefficient of the dual operator to the mixing time of the corresponding Markov chain regardless of the norm chosen. Consequently, our tighter bound on the contraction coefficient implies a tighter bound on the mixing time. We offer a result that provides an intuitive understanding of what it means to be close in a specific norm (relating the probability of any event with the probability of the same event under the stationary measure π and a ψ-Orlicz/Amemiya-norm). We then focus on L_p norms and show that asking for a bounded norm with larger p guarantees a faster decay in the probability. This is particularly relevant for exponentially decaying probabilities under π. Moreover, by exploiting the flexibility offered by Orlicz spaces, we can tackle settings where the stationary distribution is heavy-tailed, a severely under-studied setup.

2. Improved concentration bounds for MCMC methods leading to improved lower bounds on the burn-in period: by leveraging L_p-norms with large p and our results on the contraction coefficient, similar to the approach undertaken for the mixing times, we can provide improved exponential concentration bounds for MCMC methods.

3. Improved concentration bounds for sequences of Markovian random variables: we show how our results can be used to outperform existing bounds based on a change of measure technique for random variables with a Markovian dependence. In particular, we can prove exponential concentration in new settings (inaccessible to earlier approaches) and improve the rate in others.

Keywords: Markovian operators, Orlicz spaces, Contraction, MCMC, McDiarmid, Burn-in

. Extended abstract. Full version appears as [2402.11200, v3]

Acknowledgments

The authors are partially supported by the 2019 Lopez-Loreta Prize. They would also like to thank the anonymous reviewers for their careful reading of the manuscript and for providing valuable suggestions and comments, identifying a typo and suggesting meaningful references.

References

Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5, 04 2012.

Olivier Cappé, Eric Moulines, and Tobias Ryden. *Inference in Hidden Markov Models (Springer Series in Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2005.

Kai-Min Chung, Henry Lam, Zhenming Liu, and Michael Mitzenmacher. Chernoff-hoeffding bounds for Markov chains: Generalized and simplified. In *Symposium on Theoretical Aspects of Computer Science*, 2012.

Imre Csiszár. Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizitat von markoffschen ketten. *Magyar. Tud. Akad. Mat. Kutató Int. Közl,*, 8:85–108, 1963.

Pierre Del Moral, Michel Ledoux, and Laurent Miclo. On contraction properties of markov kernels. *Probability Theory and Related Fields*, 126:395–420, 2003.

Persi Diaconis and Laurent Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *The Annals of Applied Probability*, 6(3):695 – 750, 1996.

Ian H. Dinwoodie. A probability inequality for the occupation measure of a reversible Markov chain. *The Annals of Applied Probability*, 5(1):37–43, 1995.

Roland L. Dobrushin. Central limit theorem for nonstationary Markov chains. i. *Theory of Probability & Its Applications*, 1(1):65–80, 1956.

Amedeo Roberto Esposito and Marco Mondelli. Concentration without independence via information measures. *IEEE Transactions on Information Theory*, 70(6):3823–3839, 2024.

Amedeo Roberto Esposito, Michael Gastpar, and Ibrahim Issa. Generalization error bounds via Rényi-, f-divergences and maximal leakage. *IEEE Transactions on Information Theory*, 67(8): 4986–5004, 2021.

Jianqing Fan, Bai Jiang, and Qiang Sun. Hoeffding's inequality for general markov chains and its applications to statistical learning. *Journal of Machine Learning Research*, 22(139):1–35, 2021.

Charles J. Geyer. Practical Markov Chain Monte Carlo. *Statistical Science*, 7(4):473 – 483, 1992.

Walter R. Gilks, Sylvia Richardson, and David Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis, 1995.

David Wallace Gillman. *Hidden Markov chains: convergence rates and the complexity of inference*. PhD thesis, Massachusetts Institute of Technology, Boston, US, 1993.

Henryk Hudzik and Lech Maligranda. Amemiya norm equals orlicz norm in general. *Indagationes Mathematicae*, 11(4):573 – 585, 2000.

Søren Jarner and Gareth Roberts. Polynomial convergence rates of markov chains. *Annals of Applied Probability*, 12, 2000.

Søren Jarner and Gareth Roberts. Convergence of heavy-tailed monte carlo markov chain algorithms. *Scandinavian Journal of Statistics*, 34:781–815, 2007.

Carlos A. León and François Perron. Optimal Hoeffding bounds for discrete reversible Markov chains. *The Annals of Applied Probability*, 14(2):958–970, 2004.

Błażej Miasojedow. Hoeffding's inequalities for geometrically ergodic Markov chains on general state space. *Statistics & Probability Letters*, 87:115–120, 2014.

Z.D. Ren M.M. Rao. *Theory of Orlicz Spaces*. New York: M. Dekker, 1991.

Maxim Raginsky. Strong data processing inequalities and ϕ-sobolev inequalities for discrete channels. *IEEE Transactions on Information Theory*, 62(6):3355–3389, 2016.

Maxim Raginsky and Igal Sason. Concentration of measure inequalities in information theory, communications, and coding. *Foundations and Trends in Communications and Information Theory*, 10(1-2):1–246, 2013.

Cyril Roberto and Bogusław Zegarlinski. Hypercontractivity for Markov semi-groups. *Journal of Functional Analysis*, 282(12):109439, 2022.

Gareth Roberts and Jeffrey Rosenthal. Geometric Ergodicity and Hybrid Markov Chains. *Electronic Communications in Probability*, 2(none):13 – 25, 1997.

Gareth O. Roberts and Jeffrey S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 2004.

Daniel Rudolf. Explicit error bounds for Markov chain Monte Carlo. *Dissertationes Mathematicae*, 485:1–93, 2012.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

Krzystof Łatuszyński, Błażej Miasojedow, and Wojciech Niemiro. Nonasymptotic bounds on the estimation error of mcmc algorithms. *Bernoulli*, 19(5A):2033–2066, 2013.