The Journal of Open Source Software

volesti: A C++ library for sampling and volume computation on convex bodies

Apostolos Chalkis [2,4*], **Vissarion Fisikopoulos** [1,4*¶], **Marios Papachristou** [5*], and **Elias Tsigaridas**[3,4*]

1 National & Kapodistrian University of Athens, Greece **2** Quantagonia **3** Inria Paris and IMJ-PRG, Sorbonne Université **4** GeomScale **5** Cornell University ¶ Corresponding author * These authors contributed equally.

DOI: 10.21105/joss.07886

Software

- Review
- Repository
- Archive

Editor: Olexandr Konovalov

Reviewers:

- @bsun1220
- @pregnault

Submitted: 11 December 2024
Published: 11 April 2025

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Summary

Sampling from (constrained) high-dimensional distributions and volume approximation of convex bodies are fundamental operations that appear in optimization, finance, engineering, artificial intelligence, and machine learning. We present `volesti`, a C++ library that delivers efficient implementations of state-of-the-art, mainly randomized, algorithms to sample from general logarithmically concave (or log-concave) distributions. Based on these routines, we can estimate the volume of convex bodies in high dimensions, round them, and compute multidimensional integrals over them. The backbone of our library consists of Monte Carlo algorithms, which are randomized algorithms, the output of which can be incorrect with (usually very small) error probability; thus, we also provide several high-dimensional statistical tests to certify and verify the output.

The focus of `volesti` is scalability in high dimensions, that, depending on the problem at hand, could range from hundreds to thousands of dimensions. Another novelty is the ability to handle a variety of different inputs for the constrained support of the various distributions. `volesti` supports three different types of polyhedra (Ziegler, 1995), spectrahedra (Ramana & Goldman, 1999) and general non-linear convex objects.

`volesti` relies on `Eigen` library (Guennebaud et al., 2010) for linear algebra but also supports `MKL` optimizations (*Intel Math Kernel Library (Intel MKL)*, 2024). There are R (Chalkis & Fisikopoulos, 2021) and Python (Chalkis, Fisikopoulos, Tsigaridas, et al., 2023) interfaces available.

Statement of need

High-dimensional sampling from multivariate distributions with Markov Chain Monte Carlo (MCMC) algorithms is a fundamental problem with many applications in science and engineering (Genz & Bretz, 2009; Iyengar, 1988; Schellenberger & Palsson, 2009; Somerville, 1998). In particular, multivariate integration over a convex set as well as the volume approximation of convex sets have garnered significant attention from theorists and engineers over the last decades. Nevertheless, these problems are computationally hard for general dimensions (Dyer & Frieze, 1988). MCMC algorithms made remarkable progress and their use allowed us to efficiently tackle the problems of sampling and volume estimation of convex bodies in theory, by the introduction of (rigorous) theoretical guarantees (Chen et al., 2018; Lee & Vempala, 2018; Mangoubi & Vishnoi, 2019). Unfortunately, these theoretical guarantees of the MCMC algorithms do not extend in a straightforward manner to efficient implementations able to attack problems coming from real-life computations. Therefore, we witnessed the birth of efficient in practice MCMC algorithm that relax the theoretical guarantees and employ new algorithmic and statistical techniques to be amenable to efficient implementations. Remarkably, these algorithms, and the corresponding implementations, also meet the requirements for high accuracy results (Chalkis, Emiris, et al., 2023; Cousins & Vempala, 2016; Emiris & Fisikopoulos, 2014; Kook et al., 2022); however several existing published methods are available as part of propertiary packages (MATLAB) (Cousins & Vempala, 2016; Kook et al., 2022).

Our open-source package – `volesti` – offers all of the aforementioned functionality, together with the support of sampling from general log-concave densities (Chalkis, Fisikopoulos, Papachristou, et al., 2023), and uniform sampling from spectrahedra (Chalkis et al., 2022).

Our implementation: 1. supports various sampling techniques based on geometric walks; roughly speaking these are a continuous version of MCMC algorithms, such as Billard walk, Hamiltonian walk and others, 2. gives the user the ability to sample from various distributions, like uniform, exponential, Gaussian, and general log-concave densities, 3. allows to consider the distributions constrained in various convex domains, such as hypercubes, zonotopes, general polytopes (defined either as a set of linear inequalities or as a convex hull of a pointset), spectrahedra (feasible sets of semidefinite programs), and, 4. can perform volume computations, integration, and solve problems from real life applications in very high dimensions.

Impact

`volesti` has been used extensively in various research and engineering projects coauthored by the authors of this paper. In particular, for the problem of sampling the flux space of metabolic networks we were able to sample from the most complicated human metabolic network accessible today, Recon3D (Chalkis et al., 2021), used to model financial crises (Calès et al., 2023), to detect low volatility anomalies in stock markets (Bachelard et al., 2023), to introduce randomized control in asset pricing and portfolio performance evaluation (Bachelard et al., 2024)), but also to sample from (and compute the volume of) spectrahedra (Chalkis et al., 2022), the feasible regions of semidefinite programs.

Even more, `volesti` has been used by other research teams in conducting research in electric power systems (Venzke et al., 2021), for problems in probabilistic inference (Spallitta et al., 2024), to perform resource analysis on programs (Pham, 2024); but also for more theoretical and mathematical challenges, like the computation of topological invariants (Maria & Rouillé, 2021), and persistent homology (Vejdemo-Johansson & Mukherjee, 2022).

Acknowledgements

We would like to thank the contributors to the `volesti` library for their valuable contributions and feedback. MP was partially supported by a Cornell University Fellowship, a grant from the A.G. Leventis Foundation, a grant from the Gerondelis Foundation, and a LinkedIn Ph.D. Fellowship.

References

Bachelard, C., Chalkis, A., Fisikopoulos, V., & Tsigaridas, E. (2023). Randomized geometric tools for anomaly detection in stock markets. *International Conference on Artificial Intelligence and Statistics*, 9400–9416.

Bachelard, C., Chalkis, A., Fisikopoulos, V., & Tsigaridas, E. (2024). Randomized control in performance analysis and empirical asset pricing. *Ssrn Preprint*. https://doi.org/10.2139/ssrn.4744249

Calès, L., Chalkis, A., Emiris, I. Z., & Fisikopoulos, V. (2023). Practical volume approximation of high-dimensional convex bodies, applied to modeling portfolio dependencies and financial crises. *Computational Geometry*, *109*, 101916. https://doi.org/10.1016/j.comgeo.2022.101916

Chalkis, A., Emiris, I. Z., & Fisikopoulos, V. (2023). A practical algorithm for volume estimation based on billiard trajectories and simulated annealing. *ACM J. Exp. Algorithmics*, *28*. https://doi.org/10.1145/3584182

Chalkis, A., Emiris, I. Z., Fisikopoulos, V., Repouskos, P., & Tsigaridas, E. (2022). Efficient sampling in spectrahedra and volume approximation. *Linear Algebra and Its Applications*, *648*, 205–232. https://doi.org/10.1016/j.laa.2022.04.002

Chalkis, A., & Fisikopoulos, V. (2021). volesti: Volume Approximation and Sampling for Convex Polytopes in R. *The R Journal*, *13*(2), 642–660. https://doi.org/10.32614/RJ-2021-077

Chalkis, A., Fisikopoulos, V., Papachristou, M., & Tsigaridas, E. (2023). Truncated log-concave sampling for convex bodies with reflective hamiltonian monte carlo. *ACM Trans. Math. Softw.*, *49*(2). https://doi.org/10.1145/3589505

Chalkis, A., Fisikopoulos, V., Tsigaridas, E. P., & Zafeiropoulos, H. (2021). Geometric algorithms for sampling the flux space of metabolic networks. In K. Buchin & Éric Colin de Verdière (Eds.), *37th international symposium on computational geometry, SoCG 2021, june 7-11, 2021, Buffalo, NY, USA (virtual conference)* (Vol. 189, pp. 21:1–21:16). Schloss Dagstuhl - Leibniz-Zentrum für Informatik. https://doi.org/10.4230/LIPICS.SOCG.2021.21

Chalkis, A., Fisikopoulos, V., Tsigaridas, E., & Zafeiropoulos, H. (2023). Dingo: A python package for metabolic flux sampling. *bioRxiv*. https://doi.org/10.1101/2023.06.18.545486

Chen, Y., Dwivedi, R., Wainwright, M. J., & Yu, B. (2018). Fast MCMC sampling algorithms on polytopes. *Journal of Machine Learning Research*, *19*(55), 1–86. http://jmlr.org/papers/v19/18-158.html

Cousins, B., & Vempala, S. (2016). A practical volume algorithm. *Mathematical Programming Computation*, *8*(2). https://doi.org/10.1007/s12532-015-0097-z

Dyer, M., & Frieze, A. (1988). On the complexity of computing the volume of a polyhedron. *SIAM Journal on Computing*, *17*(5), 967–974. https://doi.org/10.1137/0217060

Emiris, I. Z., & Fisikopoulos, V. (2014). Practical polytope volume approximation. *ACM Transactions of Mathematical Software, 2018*, *44*(4), 38:1–38:21. https://doi.org/10.1145/3194656

Genz, A., & Bretz, F. (2009). *Computation of multivariate normal and t probabilities* (1st ed.). Springer Publishing Company, Incorporated. https://doi.org/10.1007/978-3-642-01689-9

Guennebaud, G., Jacob, B., & others. (2010). *Eigen v3*. http://eigen.tuxfamily.org.

Intel math kernel library (intel MKL). (2024). https://software.intel.com/en-us/intel-mkl.

Iyengar, S. (1988). Evaluation of normal probabilities of symmetric regions. *SIAM Journal on Scientific and Statistical Computing*, *9*(3), 418–423. https://doi.org/10.1137/0909028

Kook, Y., Lee, Y.-T., Shen, R., & Vempala, S. (2022). Sampling with riemannian hamiltonian monte carlo in a constrained space. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), *Advances in neural information processing systems* (Vol. 35, pp. 31684–31696). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2022/file/cdaa7f07b0c5a7803927d20aa717132e-Paper-Conference.pdf

Lee, Y. T., & Vempala, S. (2018). Convergence rate of Riemannian Hamiltonian Monte Carlo and faster polytope volume computation. *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, 1115–1121. https://doi.org/10.1145/3188745.3188774

Mangoubi, O., & Vishnoi, N. K. (2019). Faster polytope rounding, sampling, and volume computation via a sub-linear ball walk. *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, 1338–1357. https://doi.org/10.1109/focs.2019.00082

Maria, C., & Rouillé, O. (2021). Computation of large asymptotics of 3-manifold quantum invariants. *2021 Proceedings of the Workshop on Algorithm Engineering and Experiments (ALENEX)*, 193–206. https://doi.org/10.1137/1.9781611976472.15

Pham, L. (2024). *Hybrid resource-bound analyses of programs* [PhD thesis]. Carnegie Mellon University.

Ramana, M., & Goldman, A. (1999). Some geometric results in semidefinite programming. *Journal of Global Optimization*, *7*. https://doi.org/10.1007/BF01100204

Schellenberger, J., & Palsson, B. O. (2009). Use of randomized sampling for analysis of metabolic networks. *The Journal of Biological Chemistry*, *284 9*, 5457–5461. https://doi.org/10.1074/jbc.r800048200

Somerville, P. N. (1998). Numerical computation of multivariate normal and multivariate-t probabilities over convex regions. *Journal of Computational and Graphical Statistics*, *7*(4), 529–544. https://doi.org/10.2307/1390681

Spallitta, G., Masina, G., Morettin, P., Passerini, A., & Sebastiani, R. (2024). Enhancing SMT-based weighted model integration by structure awareness. *Artificial Intelligence*, *328*, 104067. https://doi.org/10.1016/j.artint.2024.104067

Vejdemo-Johansson, M., & Mukherjee, S. (2022). Multiple hypothesis testing with persistent homology. *Foundations of Data Science*, *4*(4), 667–705. https://doi.org/10.3934/fods.2022018

Venzke, A., Molzahn, D. K., & Chatzivasileiadis, S. (2021). Efficient creation of datasets for data-driven power system applications. *Electric Power Systems Research*, *190*, 106614. https://doi.org/10.1016/j.epsr.2020.106614

Ziegler, G. M. (1995). *Lectures on polytopes*. Springer-Verlag. https://doi.org/10.1007/978-1-4613-8431-1