The Journal of Open Source Software

`tmg-hmc`: A Python package for Exact HMC Sampling for Truncated Multivariate Gaussians with Linear and Quadratic Constraints

Erik A. Bensen [1¶] **and Mikael Kuusela** [1]

1 Department of Statistics and Data Science, Carnegie Mellon University ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 31 December 2025
Published: unpublished

Summary

Markov Chain Monte Carlo is a cornerstone of statistical methods that allows us to approximate intractable quantities by sampling from complex, multivariate probability distributions. One important class of distributions is constrained distributions. We present `tmg-hmc`: a Python implementation of the Exact Hamiltonian Monte Carlo for Truncated Multivariate Gaussians with linear and quadratic inequality constraints introduced in Pakman & Paninski (2014). This method leverages the high-dimensional scalability and good mixing properties of Hamiltonian Monte Carlo while maintaining speed and simplicity since the Hamiltonian equations for a truncated Gaussian distribution are analytically solvable. This means that the sampler always accepts the sampled value and there are no tunable parameters. The original authors created an R implementation `tmg` and a Matlab implementation `hmc_tmg`. Both of these implementations are no longer maintained and the R package was archived from CRAN in 2021. Bertolacci et al. (2024) partially implements the exact HMC in R, however, this implementation is limited to only linear constraints. There are also two R packages, `VeccTMVN` and `nntmvn`, that can sample truncated multivariate Gaussians. However, these implementations are both approximate and limited to linear box constraints. To the best of our knowledge, `tmg-hmc` is the only existing Python implementation of Exact HMC. Additionally, we expand our implementation by including sparse matrix operations for sparse constraint handling and optional GPU acceleration for high-dimensional problems such as truncated Gaussian processes. Finally, we accelerate the quadratic constraint hit-time calculation by using a speed-optimized C++ implementation that can be called from Python.

Statement of need

Many statistical models of real-world phenomena require the computation of intractable integrals over complex, multivariate probability distributions. Markov Chain Monte Carlo is a foundational statistical method that allows can be used to approximate these quantities from samples (Robert & Casella, 1999). This has allowed for significant progress in statistical modeling in many areas of applied statistics and machine learning (Gelman et al., 2013). One important class of distributions that arises due to parameter or data constraints are truncated distributions (Gelfand et al., 1992; Stanley et al., 2025; Swiler et al., 2020).

Pakman & Paninski (2014) consider sampling a d-dimensional Gaussian $X \sim N(\mu, \Sigma)$ that is truncated with m inequality constraints of the form

$$Q_j(X) \geq 0, \qquad j = 1, \dots, m$$

, where $Q_j(X)$ is a product of linear and quadratic polynomials. As discussed by Pakman and Paninski, this type of a distribution is a critical component of a vast array of statistical models including the probit and Tobit models (Albert & Chib, 1993; Tobin, 1958), the dichotomized Gaussian model (Cox & Wermuth, 2002; Emrich & Piedmonte, 1991), stochastic integrate-and-fire neural models (Paninski et al., 2003), Bayesian isotonic regression (Neelon & Dunson, 2004), and the Bayesian bridge model (Polson et al., 2014).

More recently, distributions of this form have been used for learning partially censored Gaussian processes (Cao & Katzfuss, 2025). The Exact HMC algorithm was used to implement physics-informed constraints for fields governing CO2 flux in the WOMBAT v2.0 hierarchical flux-inversion framework for inferring changes to the global carbon cycle (Bertolacci et al., 2024). Additionally, sampling constrained multivariate normals is relevant to a growing range of literature on constrained Gaussian processes (Agrell, 2019; Bachoc et al., 2019, 2022; Da Veiga & Marrel, 2012; López-Lopera et al., 2018; Maatouk, 2017; Swiler et al., 2020). In particular, in our ongoing research, we are using `tmg-hmc` to sample random transport maps given by the gradient of 2d convex Gaussian processes. We do this by approximating the convex GP by imposing quadratic convex inequality constraints on a discrete spatial grid. Figure 1 shows an example of such a transport map sampled using `tmg-hmc`.

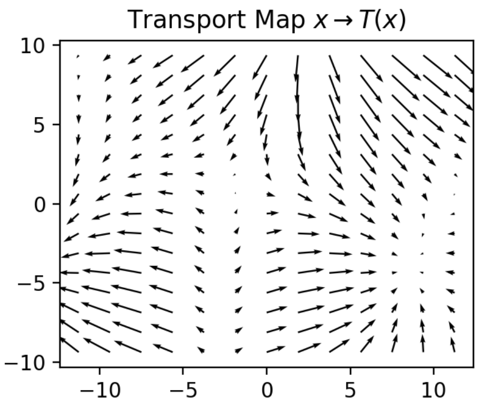

Figure 1: Sample of a random transport map defined as the gradient of a 2d convex Gaussian process.

Exact HMC is not the only method for sampling distributions of this family. Two main alternatives include classical Hamiltonian Monte Carlo (Duane et al., 1987; Neal, 2011)and Gibbs sampling with the Hit-and-Run Algorithm (Chen & Deely, 1992). HMC is a fast-mixing algorithm that is robust to high numbers of dimensions. However, generally speaking, it requires integrating equations of motion and using a Metropolis accept-reject step to account for numerical integration error. The numerical integration also comes with its own tunable hyperparameters that must be adjusted to balance exploration of the state space with a high acceptance probability (Hoffman et al., 2014). On the other hand, Gibbs sampling is a simpler method with no hyperparameters that always accepts samples, however, it can be slow to mix, particularly when constraints impose high correlation between variables. Since the constrained Gaussian HMC trajectories are analytically computable Exact HMC enables the best of both options, the good mixing and high-dimensional capabilities of classical HMC with the always accepting and no hyperparameter properties of the Gibbs sampler. See the original manuscript Pakman & Paninski (2014) for a more detailed discussion of the differences between these methods. Some other alternative methods for sampling truncated multivariate Gaussian distributions include the R packages `VeccTMVN` (Cao & Katzfuss, 2024) and `nntmvn` (Cao & Katzfuss, 2025) which use Vecchia and nearest-neighbor approximations, respectively, to sample from a truncated Gaussian. However, these methods are both approximate and limited to sampling Gaussians with linear box constraints. `tmg-hmc` is developed as a flexible, user friendly and well tested Python package so that anyone can leverage the benefits of Exact HMC without needing to dwell on the technical details.

Basic Usage

tmg-hmc operates predominantly through the TMGSampler class where a user specifies the untruncated distribution, adds constraints and then samples from the truncated distribution as illustrated in the example code below. All of the HMC trajectories and constraint hit-time solutions are handled automatically behind the scenes by the class internals.

```
import numpy as np
from tmg_hmc import TMGSampler

# Set up untruncated distribution parameters
mu = np.array([0., 1.]).reshape(-1,1)
sigma = np.array([[1., 0.6],[0.6, 1.]])
sampler = TMGSampler(mu, sigma)

# Add constraints
# Second coordinate positive
f_positivity = np.array([0., 1.]).reshape(-1,1)
c_positivity = 0
sampler.add_constraint(f=f_positivity, c=c_positivity)

# Bounded outside of unit circle
A_unit = np.eye(2)
c_unit = -1
sampler.add_constraint(A=A_unit, c=c_unit)

# Run the exact HMC sampling algorithm
x0 = np.array([2., 1.]).reshape(-1,1)
samples = sampler.sample(x0, n_samples=1000, burn_in=100)
```

Acknowledgments

This work was partially supported by the U.S. National Science Foundation under Grant No. DMS-2310632.

References

Agrell, C. (2019). Gaussian Processes with Linear Operator Inequality Constraints. *Journal of Machine Learning Research*, *20*(135), 1–36.

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*(422), 669–679.

Bachoc, F., Lagnoux, A., & López-Lopera, A. F. (2019). Maximum likelihood estimation for Gaussian processes under inequality constraints. *Electronic Journal of Statistics*, *13*(2), 2921–2969.

Bachoc, F., López-Lopera, A. F., & Roustant, O. (2022). Sequential Construction and Dimension Reduction of Gaussian Processes Under Inequality Constraints. *SIAM Journal on Mathematics of Data Science*, *4*(2), 772–800.

Bertolacci, M., Zammit-Mangion, A., Schuh, A., Bukosa, B., Fisher, J. A., Cao, Y., Kaushik, A., & Cressie, N. (2024). Inferring changes to the global carbon cycle with WOMBAT v2.0, a hierarchical flux-inversion framework. *The Annals of Applied Statistics*, *18*(1), 303–327.

Cao, J., & Katzfuss, M. (2024). *Scalable Sampling of Truncated Multivariate Normals Using*

Sequential Nearest-Neighbor Approximation. Preprint arXiv:2406.17307.

Cao, J., & Katzfuss, M. (2025). *Linear-Cost Vecchia Approximation of Multivariate Normal Probabilities*. Preprint arXiv:2311.09426.

Chen, M.-H., & Deely, J. (1992). *Application of a new Gibbs hit-and-run sampler to a constrained linear multiple regression problem*. Technical Report 92-21, Purdue University, Center for Statistical Decision Sciences; Department of Statistics.

Cox, D. R., & Wermuth, N. (2002). On some models for multivariate binary variables parallel in complexity with the multivariate Gaussian distribution. *Biometrika*, 462–469.

Da Veiga, S., & Marrel, A. (2012). Gaussian process modeling with inequality constraints. *Annales de La Faculté Des Sciences de Toulouse : Mathématiques*, *21*(3), 529–555.

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, *195*(2), 216–222.

Emrich, L. J., & Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variates. *The American Statistician*, *45*(4), 302–304.

Gelfand, A. E., Smith, A. F., & Lee, T.-M. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association*, *87*(418), 523–532.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC Press.

Hoffman, M. D., Gelman, A., & others. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, *15*(1), 1593–1623.

López-Lopera, A. F., Bachoc, F., Durrande, N., & Roustant, O. (2018). Finite-Dimensional Gaussian Approximation with Linear Inequality Constraints. *SIAM/ASA Journal on Uncertainty Quantification*, *6*(3), 1224–1255.

Maatouk, H. (2017). *Finite-dimensional approximation of Gaussian processes with inequality constraints*. Preprint arXiv:1706.02178.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, *2*(11), 2.

Neelon, B., & Dunson, D. B. (2004). Bayesian isotonic regression and trend analysis. *Biometrics*, *60*(2), 398–406.

Pakman, A., & Paninski, L. (2014). Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *Journal of Computational and Graphical Statistics*, *23*(2), 518–542.

Paninski, L., Simoncelli, E., & Pillow, J. (2003). Maximum likelihood estimation of a stochastic integrate-and-fire neural model. *Advances in Neural Information Processing Systems*, *16*.

Polson, N. G., Scott, J. G., & Windle, J. (2014). The bayesian bridge. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *76*(4), 713–733.

Robert, C. P., & Casella, G. (1999). *Monte Carlo statistical methods* (Vol. 2). Springer.

Stanley, M., Batlle, P., Patil, P., Owhadi, H., & Kuusela, M. (2025). Confidence intervals for functionals in constrained inverse problems via data-adaptive sampling-based calibration. *Preprint arXiv:2502.02674*.

Swiler, L. P., Gulian, M., Frankel, A. L., Safta, C., & Jakeman, J. D. (2020). A Survey of Constrained Gaussian Process Regression: Approaches and Implementation Challenges. *Journal of Machine Learning for Modeling and Computing*, *1*(2).

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: Journal of the Econometric Society*, 24–36.