Stein's Lemma for the Reparameterization Trick with Gaussian Mixtures

Wu Lin [1] **Mohammad Emtiyaz Khan** [2] **Mark Schmidt** [1]

Abstract

Stein's method and Stein's lemma (Stein, 1973; 1981) are both powerful tools for statistical applications, and have had a significant impact in machine learning. Previous applications of Stein's lemma either required strong technical assumptions or were limited to Gaussian distributions with restricted covariance structures. For example, Park et al. (2012) and Brown et al. (2006) use Stein's lemma for De Bruijn's identity and the heat equation respectively, but require strong assumptions, such as diagonal covariance structure and twice continuous differentiability to simplify the proof. Bonnet (1964) and Price (1958) use characteristic functions to derive extensions for arbitrary covariance structure, but their proofs do not easily extend to Gaussian mixtures. Adcock (2007); Adcock & Shutes (2012); Landsman (2006); Landsman & Nešlehová (2008) extend Stein's lemma to a class of Gaussian mixtures, but their methods do not naturally lead to a second-order estimation, such as Price's theorem (Price, 1958), which is an unbiased, low-variance estimator (Salimans & Knowles, 2013; Erdogdu, 2015; Khan et al., 2017). Other works such as Fan et al. (2015); Erdogdu (2015); Rezende et al. (2014) apply Stein's lemma to Gaussian approximations, where the authors use integration by parts to extend the lemma, but the specific technical conditions in their derivations are restrictive.

In this work, we extend Stein's lemma to flexible Gaussian-mixture distributions under weak assumptions. Our generalization enables us to establish a connection between Stein's lemma and the reparamterization trick to derive gradients of expectations of a large class of functions. Using this connection, we can derive many new reparameterizable gradient-identities that goes beyond the reach of existing works under weak assumptions. For example, we give gradient identities when expectation is taken with respect to Student's t-distribution, skew Gaussian, exponentially modified Gaussian, and normal inverse Gaussian. Finally, we show applications of these identities to approximate the posterior distribution of complex models using variational inference.

[1]University of British Columbia, Vancouver, Canada. [2]RIKEN Center for Advanced Intelligence Project, Tokyo, Japan. Correspondence to: Wu Lin <wlin2018@cs.ubc.ca>.

Preliminary work. Under review by the ICML Workshop on Stein's Method, 2019. Do not distribute.

References

Adcock, C. Extensions of stein's lemma for the skew-normal distribution. *Communications in Statistics-Theory and Methods*, 36(9):1661–1671, 2007.

Adcock, C. and Shutes, K. On the multivariate extended skew-normal, normal-exponential, and normal-gamma distributions. *Journal of Statistical Theory and Practice*, 6(4):636–664, 2012.

Bonnet, G. Transformations des signaux aléatoires a travers les systemes non linéaires sans mémoire. In *Annales des Télécommunications*, volume 19, pp. 203–220. Springer, 1964.

Brown, L., DasGupta, A., Haff, L. R., and Strawderman, W. E. The heat equation and stein's identity: Connections, applications. *Journal of Statistical Planning and Inference*, 136(7):2254–2278, 2006.

Erdogdu, M. A. Newton-stein method: a second order method for glms via stein's lemma. In *Advances in Neural Information Processing Systems*, pp. 1216–1224, 2015.

Fan, K., Wang, Z., Beck, J., Kwok, J., and Heller, K. A. Fast second order stochastic backpropagation for variational inference. In *Advances in Neural Information Processing Systems*, pp. 1387–1395, 2015.

Khan, M. E., Lin, W., Tangkaratt, V., Liu, Z., and Nielsen, D. Variational adaptive-newton method for explorative learning. *arXiv preprint arXiv:1711.05560*, 2017.

Landsman, Z. On the generalization of stein's lemma for elliptical class of distributions. *Statistics & probability letters*, 76(10):1012–1016, 2006.

Landsman, Z. and Nešlehová, J. Stein's lemma for elliptical random vectors. *Journal of Multivariate Analysis*, 99(5): 912–927, 2008.

Park, S., Serpedin, E., and Qaraqe, K. On the equivalence between stein and de bruijn identities. *IEEE Transactions on Information Theory*, 58(12):7045–7067, 2012.

Price, R. A useful theorem for nonlinear devices having gaussian inputs. *IRE Transactions on Information Theory*, 4(2):69–72, 1958.

Rezende, D. J., Mohamed, S., and Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*, 2014.

Salimans, T. and Knowles, D. Fixed-form variational posterior approximation through stochastic linear regression. *Bayesian Analysis*, 8(4):837–882, 2013.

Stein, C. Estimation of the mean of a multivariate normal distribution. *Proc. Prague Sympos. Asymptotic Statistics*, 1973.

Stein, C. M. Estimation of the mean of a multivariate normal distribution. *The annals of Statistics*, pp. 1135–1151, 1981.