

Proceeding Paper

Adaptive Importance Sampling for Equivariant Group-Convolution Computation [†]

Pierre-Yves Lagrave ^{1,*}  and Frédéric Barbaresco ² ¹ Thales Research and Technology, 91767 Palaiseau, France² Thales Land and Air Systems, 92190 Meudon, France

* Correspondence: pierre-yves.lagrave@thalesgroup.com

[†] Presented at the 41st International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Paris, France, 18–22 July 2022.

Abstract: This paper introduces an adaptive importance sampling scheme for the computation of group-based convolutions, a key step in the implementation of equivariant neural networks. By leveraging information geometry to define the parameters update rule for inferring the optimal sampling distribution, we show promising results for our approach by working with the two-dimensional rotation group $SO(2)$ and von Mises distributions. Finally, we position our AIS scheme with respect to quantum algorithms for computing Monte Carlo estimations.

Keywords: group convolution; adaptive importance sampling; equivariant neural networks; natural gradient; information geometric optimization; quantum algorithms



Citation: Lagrave, P.-Y.; Barbaresco, F. Adaptive Importance Sampling for Equivariant Group-Convolution Computation. *Phys. Sci. Forum* **2022**, *5*, 17. <https://doi.org/10.3390/psf2022005017>

Academic Editors: Ali Mohammad-Djafari, Frank Nielsen and Martino Trassinelli

Published: 5 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction and Motivations

Geometric deep learning [1] is an emerging field receiving more and more traction because of its successful application to a wide range of domains [2–4]. In this context, equivariant neural networks (ENN) [5] have been shown to be superior to conventional deep learning approaches from both accuracy and robustness standpoints and appear as a natural alternative to data augmentation techniques [6,7] to achieve geometrical robustness. One key bottleneck for scaling ENN to industrial applications lies with the numerical computation of the associated equivariant operators. More precisely, two main approaches have been used in the literature, namely a Monte Carlo sampling method [2] (which can be made exhaustive for small finite groups) and a generalized Fourier-based method [4,8,9]. However, these approaches suffer from scalability issues as the complexity of the underlying group increases (e.g., handling non-compact groups such as $SU(1,1)$ or large finite groups such as the symmetric group S_n is challenging). Even for groups such as $SO(2)$ for which previous works on the use of spherical harmonics can be leveraged on, the efficient computation of a reliable estimate of the convolution remains a challenge (convergence). In this context, the authors of [10] have proposed an efficient method for building adequate kernel functions to be used within steerable neural networks [11] by leveraging on the knowledge of infinitesimal generators of the considered Lie group and on a Krylov approach for solving the linear constraints. We propose in this paper to cover the specific case of group-convolutional neural networks (G-CNN) [2,12], which in particular, rely on the computation of group-based convolution operators. By leveraging on information geometry as proposed in [13] for quantile estimation, we introduce here an adaptive importance sampling (AIS) variance reduction method based on information geometric optimization [14] to improve the convergence of Monte Carlo estimators for the numerical computation of group-based convolution feature maps, as used in several recent works [2,9,15]. We illustrate our approach on the two-dimensional rotation group $SO(2)$ by regularizing with von Mises distributions [16], a set-up for which the Fisher information metric [17] can be computed using closed form formulas.

Finally, we shed some light on the benefits of working toward a quantum version of our proposed AIS scheme in order to reach a quadratic speed-up [18]. Improving quantum Monte Carlo integration schemes is indeed a very active topic of research [19], mainly driven by applications within the financial industry [20]. Benchmarking with group-Fourier transform-based approaches, such as [21], which are more theoretically involved but with a promise of an exponential speed-up, will be of particular interest in this context.

2. Group Convolution and Expectation

We consider in the following a compact group G with corresponding Haar measure μ^G . As $\mu^G(G) < \infty$, we can choose μ^G so that $\int_G d\mu^G = 1$ by using an adequate normalization.

We are interested in evaluating the group-based convolution operator ψ^G defined below for functionals $f, k : G \rightarrow \mathbb{R}$ and $g \in G$:

$$\psi^G(g) = \int_G k(h^{-1}g)f(h)d\mu^G(h) \tag{1}$$

Using a probabilistic interpretation of (1), we can write

$$\psi^G(g) = \mathbb{E}^{\mu^G} \left[k(H^{-1}g)f(H) \right] \tag{2}$$

where H is a G -valued random variable distributed according to μ^G . The convolution can therefore be estimated with a Monte Carlo method by using the following estimator

$$\tilde{\psi}_n^G(g) = \frac{1}{n} \sum_{i=1}^n k(h_i^{-1}g)f(h_i) \tag{3}$$

where $h_i \sim \mu^G$ and for which the efficiency could be improved through variance reduction techniques [22]. By anchoring in [13], we describe in the following an adaptive importance sampling approach for the computation of (1). Similar ideas were also used in [23] for financial applications.

3. Adaptive Importance Sampling

We consider in the following a set Φ_Θ of parametric probability density functions on G , where Θ represents the parameters space. Each density $\phi_\theta \in \Phi_\Theta$ is assumed to be absolutely continuous with respect to the Haar measure μ^G of the group G , so that the corresponding probability measure can be written as $d\mu_\theta = \phi_\theta d\mu^G$ and the Radon–Nikodym derivative $\omega_\theta = \frac{d\mu^G}{d\mu_\theta}$ can be considered.

Using the conventional importance sampling approach, we can then write:

$$\psi^G(g) = \int_G k(h^{-1}g)f(h) \left(\frac{d\mu^G}{d\mu_\theta} \right) (h) d\mu_\theta(h) \tag{4}$$

$$= \mathbb{E}^{\mu_\theta} \left[\omega_\theta(H)k(H^{-1}g)f(H) \right] \tag{5}$$

The idea is then to choose a measure μ_{θ^*} for which θ^* minimizes the variance $v^{k,f,g}$ of the random variable $k(H^{-1}g)\omega_\theta(H)f(H)$, which can be written as

$$v^{k,f,g}(\theta) = \mathbb{E}^{\mu_\theta} \left[\left(\omega_\theta(H)k(H^{-1}g)f(H) \right)^2 \right] - \psi^G(g)^2 \tag{6}$$

$$= m_2^{k,f,g}(\theta) - \psi^G(g)^2 \tag{7}$$

where

$$m_2^{k,f,g}(\theta) = \mathbb{E}^{\mu_\theta} \left[\left(k(H^{-1}g)\omega_\theta(H)f(H) \right)^2 \right] \tag{8}$$

$$= \mathbb{E}^{\mu^G} \left[\omega_\theta(H) \left(k(H^{-1}g)f(H) \right)^2 \right] \tag{9}$$

3.1. Monte Carlo Estimator and Convergence

We assume that we can construct a sequence of parameters $(\theta_i)_{i=0}^{n-1}$, together with realizations $(h_i)_{i=1}^n$ of the random variables $(H_i)_{i=1}^n$ such that $H_i \sim \mu_{\theta_{i-1}}$ and that $\theta_n \rightarrow \theta^* \in \Theta$ as $n \rightarrow \infty$. We can then consider the following Monte Carlo estimator:

$$\hat{\psi}_n^G(g) = \frac{1}{n} \sum_{i=1}^n \omega_{\theta_{i-1}}(h_i) k(h_i^{-1}g) f(h_i) \tag{10}$$

Under usual integrability conditions, Theorem 3.1 of [13] states that $\hat{\psi}_n^G(g) \rightarrow \psi^G(g)$ almost surely as $n \rightarrow \infty$. Furthermore, we have the following distributional convergence result,

$$\sqrt{n}(\hat{\psi}_n^G(g) - \psi^G(g)) \rightarrow \mathcal{N}(0, v^{k,f,g}(\theta^*)) \tag{11}$$

where $\mathcal{N}(0, \sigma^2)$ refers to the Gaussian distribution with 0 mean and variance σ^2 .

3.2. Natural Gradient Descent

We now discuss how to build the sequence of parameters $(\theta_i)_{i=0}^{n-1}$ and corresponding realizations $(h_i)_{i=1}^n$ as introduced in Section 3.1, reminding ourselves that we have

$$\theta^* = \arg \min_{\theta \in \Theta} v^{k,f,g}(\theta) = \arg \min_{\theta \in \Theta} m_2^{k,f,g}(\theta) \tag{12}$$

Assuming that the parameter space $\Theta \subseteq \mathbb{R}^m$ is a smooth manifold, we can consider the Fisher information metric g on the density space Φ_Θ , which is defined as it follows [17]:

$$g_{ij} = \mathbb{E}^{\mu_\theta} \left[\frac{\partial \log \phi_\theta}{\partial \theta_i} \frac{\partial \log \phi_\theta}{\partial \theta_j} \right] \tag{13}$$

We then propose using a natural gradient descent strategy to minimize the quantity $m_2^{k,f,g}$, namely

$$\theta_{k+1} = \theta_k - \alpha_k F_k^{-1} \nabla (m_2^{k,f,g}(\theta_k)) \tag{14}$$

where F_k is the Fisher information matrix, i.e., the representation of the Fisher metric as a $m \times m$ matrix and $\alpha_k \in \mathbb{R}_+^*$. Assuming that the considered functions are smooth enough, it is possible to write:

$$\nabla (m_2^{k,f,g}(\theta)) = \mathbb{E}^{\mu^G} \left[\nabla \omega_\theta(H) \left(k(H^{-1}g) f(H) \right)^2 \right] \tag{15}$$

$$= -\mathbb{E}^{\mu^G} \left[\omega_\theta(H) \nabla \log \phi_\theta(H) \left(k(H^{-1}g) f(H) \right)^2 \right] \tag{16}$$

$$= -\mathbb{E}^{\mu_\theta} \left[\nabla \log \phi_\theta(H) \left(\omega_\theta(H) k(H^{-1}g) f(H) \right)^2 \right] \tag{17}$$

Using a stochastic approximation scheme such as the Robbins–Monro algorithm [24] then leads to consider the following update rule,

$$\theta_{k+1} = \theta_k + \alpha_k F_k^{-1} \Lambda(H_{k+1}, \theta_k) \tag{18}$$

where $\Lambda(H, \theta) = \nabla \log \phi_\theta(H) (\omega_\theta(H) k(H^{-1}g) f(H))^2$, $H_k \sim \mu_{\theta_{k-1}}$ and the sequence of numbers α_k is such that $\sum_k \alpha_k = \infty$ and $\sum_k \alpha_k^2 < \infty$.

3.3. About IGO Algorithms

Information geometric optimization (IGO) algorithms are introduced in [14] as a unified framework to solve black-box optimization problems. IGO algorithms can be seen

as performing an estimation of a distribution over the considered search space \mathcal{X} leading to small values of the target function Q when sampling according to it. More precisely, the idea is to maintain at each iteration t a parametric probability distribution P_{λ_t} on the search space \mathcal{X} , for $\lambda_t \in \Lambda \subseteq \mathbb{R}^p$ and to have the value λ_t evolve over time as to shift P_{λ_t} toward giving more weight to points $x \in \mathcal{X}$ associated with a lower value of Q .

The IGO algorithms described [14] first transfer the function Q from \mathcal{X} to Λ by using an adaptive quantile-based approach and then applying a natural gradient descent by leveraging on the Fisher information metric of the considered statistical model. The scheme described in the Definition 5 of [14] defines the following the update rule for the parameter λ_t :

$$\lambda_{t+dt} = \lambda_t + dt \times I^{-1}(\lambda_t) \sum_{i=1}^N \omega_i \left. \frac{\partial \ln P_{\lambda}(x_{i:N})}{\partial \lambda} \right|_{\lambda=\lambda_t} \tag{19}$$

where I is the fisher matrix of the model, x_1, \dots, x_N are N samples drawn according to P_{λ_t} at step t , $x_{i:N}$ denotes the sample point ranked i^{th} according to Q (i.e., $Q(x_{1:N}) < \dots < Q(x_{N:N})$) and $\omega_i = \frac{1}{N} \omega\left(\frac{i-1}{2}\right)$, with $\omega(q) = 1_{q < q_0}$ a quantile-based selection function of threshold q_0 .

IGO algorithms could therefore be used in our context by setting $Q = m_2^{k,f,g}(\theta)$ and $\mathcal{X} = \Theta$ to infer the optimal value $\theta^* \in \Theta$. Implementing the update rule (19) requires a priori a large number of evaluations of the term $Q = m_2^{k,f,g}(\theta)$ to derive the sorted samples $x_{i:N}$, making this approach generally not well suited to our context.

4. Application to SO(2)-Convolutions

We give here an application of our AIS approach for the computation of SO(2)-convolutions by using von Mises densities [16] for the weighting. This type of computation is in particular relevant when working with SE(2)-ENN by exploiting the semi-direct product structure $SE(2) = \mathbb{R}^2 \ltimes SO(2)$, as performed in [3].

4.1. Fisher Information Metric

The Haar measure on SO(2) is simply the Lebesgue measure on the unit circle $[0, 2\pi]$ and will be denoted $d\alpha$ before normalization. The convolution operator (1) is therefore applied to functionals defined on $[0, 2\pi]$, so that we are interested in evaluating the following quantity for $\alpha_0 \in [0, 2\pi]$:

$$\psi^{SO(2)}(\alpha_0) = \frac{1}{2\pi} \int_0^{2\pi} k(\alpha_0 - \alpha) f(\alpha) d\alpha \tag{20}$$

We consider in the following a family of von Mises densities ϕ_θ on $[0, 2\pi]$, for $\theta = (\mu, \kappa)$, and for which $\forall \alpha \in [0, 2\pi]$,

$$\phi_\theta(\alpha) = \frac{e^{\kappa \cos(\alpha - \mu)}}{2\pi I_0(\kappa)} \tag{21}$$

where for $n \in \mathbb{N}$, $I_n(\kappa) = \frac{1}{\pi} \int_0^\pi e^{\kappa \cos(\theta)} \cos(n\theta) d\theta$ refers to modified Bessel functions. Denoting in the following $\ell_{\mu,\kappa} = \log \phi_\theta$, we have $\partial_\mu \ell_{\mu,\kappa}(\alpha) = \kappa \sin(\alpha - \mu)$ and $\partial_\kappa \ell_{\mu,\kappa}(\alpha) = \cos(\alpha - \mu) - \frac{I'_0(\kappa)}{I_0(\kappa)}$ and it is now possible to evaluate the metric tensor g . More precisely, we easily obtain:

$$g_{\mu,\mu} = \mathbb{E}^{\mu_\theta} \left[(\partial_\mu \ell_{\mu,\kappa}(\alpha))^2 \right] \tag{22}$$

$$= \int_0^{2\pi} \kappa^2 \sin^2(\alpha - \mu) \phi_\theta(\alpha) d\alpha \tag{23}$$

$$= \kappa \frac{I_1(\kappa)}{I_0(\kappa)} \tag{24}$$

$$g_{\kappa,\kappa} = \mathbb{E}^{\mu_\theta} \left[(\partial_\kappa \ell_{\mu,\kappa}(\alpha))^2 \right] \tag{25}$$

$$= \int_0^{2\pi} \left(\cos(\alpha - \mu) - \frac{I'_0(\kappa)}{I_0(\kappa)} \right)^2 \phi_\theta(\alpha) d\alpha \tag{26}$$

$$= 1 - \frac{1}{\kappa} \frac{I_1(\kappa)}{I_0(\kappa)} - 2 \frac{I'_0(\kappa) I_1(\kappa)}{I_0(\kappa)^2} + \frac{I'_0(\kappa)^2}{I_0(\kappa)^2} \tag{27}$$

$$g_{\mu,\kappa} = \mathbb{E}^{\mu_\theta} \left[\partial_\mu \ell_{\mu,\kappa}(\alpha) \partial_\kappa \ell_{\mu,\kappa}(\alpha) \right] \tag{28}$$

$$= \int_0^{2\pi} \kappa \sin(\alpha - \mu) \left(\cos(\alpha - \mu) - \frac{I'_0(\kappa)}{I_0(\kappa)} \right) \phi_\theta(\alpha) d\alpha \tag{29}$$

$$= 0 \tag{30}$$

The inverse of the Fisher information matrix in the scheme (18) can therefore be computed with closed-form formulas, making its evaluation very efficient for the considered case.

4.2. Numerical Experiments

To numerically validate our approach, we have considered von Mises type feature functions $f_{\kappa_0, \alpha_0} : \alpha \rightarrow e^{\kappa_0 \cos(\alpha_0 - \mu)}$ and kernel functions $k : [0, 2\pi] \rightarrow \mathbb{R}$ modeled as small fully connected neural networks with one hidden layer of 128 neurons with ReLU activation and uniform random weights initialization. To run our testing, we have used $\kappa_0 = 3$ and $\mu_0 = \frac{\pi}{2}$.

Figure 1 shows the comparison between the results obtained with the estimator (10) using the adaptive importance sampling scheme and those obtained with the conventional estimator (3). We can in particular see that the adaptive importance sampling scheme converges faster to the theoretical value (here computed by using (3) with $n = 50,000$ and displayed in black in Figure 1), while providing much narrower confidence intervals (because of lower variance) than the conventional Monte Carlo estimator. Figure 2 shows the evolution of the parameter $\theta = (\mu, \kappa)$ as we iterate through the update rule (18), from which we can also observe a fast convergence.

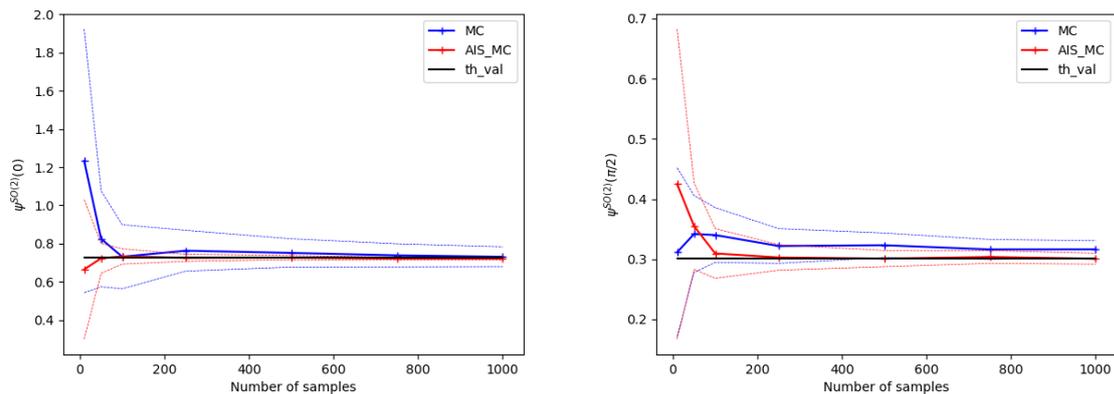


Figure 1. Comparison between the convergence of the AIS scheme with von Mises densities (red) and the traditional Monte Carlo approach (blue) by representing the evolution of the estimated convolution $\psi^{SO(2)}$ at $\alpha_0 = 0$ (left) and $\alpha_0 = \pi/2$ (right), as a function of the number n of simulated samples.

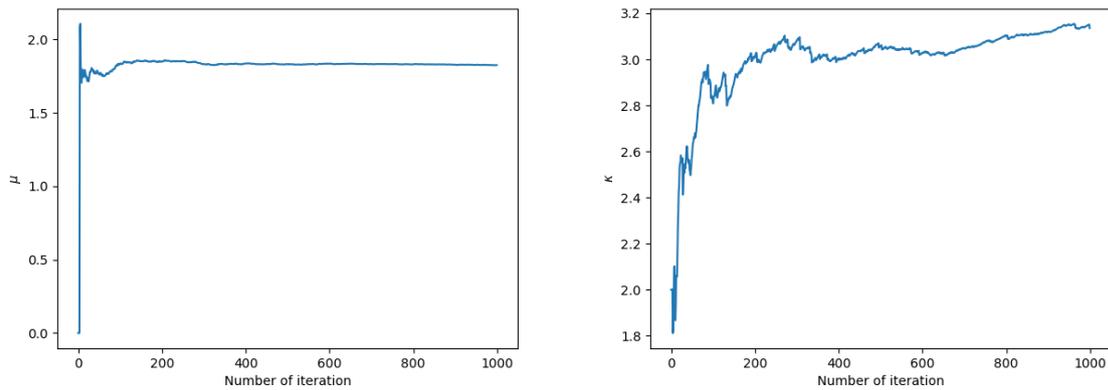


Figure 2. Evolution of the components of the parameter $\theta = (\mu, \kappa)$ as updated according to (18) when estimating $\psi^{SO(2)}(0)$ with the AIS scheme with von Mises densities.

4.3. Extension to SO(3)-Convolutions

Generalizing the above results to cover SO(3)-convolutions is of particular interest when using ENN for processing spherical data such as fish-eye images [4,25]. The Fisher-Bingham distribution [26], also known as the Kent distribution, can be leveraged in this context. More precisely, we have in this case, for $x \in \mathbb{S}_2$ (the 2D-sphere in \mathbb{R}^3):

$$\phi_{\theta}(x) = c(\kappa, \beta)^{-1} \exp\left(\kappa \gamma_1 \cdot x + \beta \left[(\gamma_2 \cdot x)^2 - (\gamma_3 \cdot x)^2\right]\right) \tag{31}$$

where γ_i for $i = 1, 2, 3$ are vectors of \mathbb{R}^3 so that the 3×3 matrix $\Gamma = [\gamma_1, \gamma_2, \gamma_3]$ is orthogonal and $c(\kappa, \beta)$ is a normalizing constant.

Although we defer to further work the details of the derivation of the corresponding AIS estimator (10), we illustrate on Figure 3 that SO(3)-convolutions could also benefit from variance reduction methods by using a simple quasi-Monte Carlo scheme [27] with a three-dimensional Sobol sequence [28].

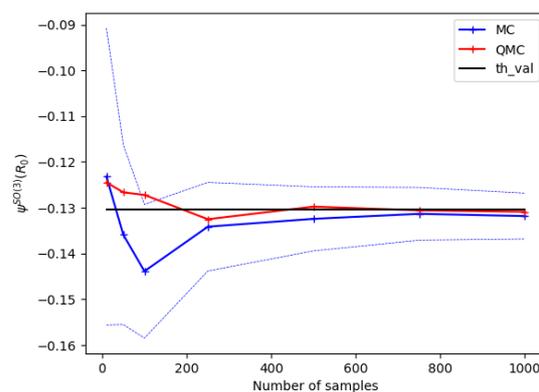


Figure 3. Convergence comparison for a SO(3)-convolution computation between the classical Monte Carlo estimate (3) with numpy random numbers generator (blue) and the corresponding quasi-Monte Carlo scheme leveraging on a Sobol sequence (red).

5. Monte Carlo Methods in the Quantum Set-Up

Monte Carlo computations can generally benefit from a quadratic speed-up in a quantum computing set-up [18] and improving quantum Monte Carlo integration schemes is a very active topic of research [19], mainly driven by applications within the financial industry [20,29].

A similar speed-up can therefore be expected in our context by estimating (1) by leveraging on the quantum amplitude estimation (QAE) algorithm [30]. For $g \in G$, we denote $\phi_g^{f,k} : G \rightarrow \mathbb{R}$ the function such that $\forall h \in G$, $\phi_g^{f,k}(h) = k(h^{-1}g)f(h)$.

We first construct the operator \mathcal{U}_{μ^G} to load a discretized version of μ^G so that $\mathcal{U}_{\mu^G}|0\rangle = \sum_{h \in G_\epsilon} \sqrt{p(h)}|h\rangle$, with $p(h) = \int_{g \in \mathcal{B}(h, r_\epsilon)} \mu^G(g)$, $\mathcal{B}(x, r)$ the ball of radius $r > 0$ centered in $x \in G$ and G_ϵ a discrete subset of G . We scale $\phi_g^{f,k}$ to $\tilde{\phi}_g^{f,k} : G \rightarrow [0, 1]$ and build another unitary operator \mathcal{U}_ϕ to compute and load the values of $\phi_g^{f,k}$ taken on G_ϵ , that we defined by $\mathcal{U}_\phi|h\rangle = \sqrt{1 - \tilde{\phi}_g^{f,k}(h)}|h\rangle|0\rangle + \sqrt{\tilde{\phi}_g^{f,k}(h)}|h\rangle|1\rangle$. Using the QAE algorithm on $\mathcal{U}_\phi \mathcal{U}_{\mu^G}$ gives us access to an estimate of (1) after proper rescaling, with a precision of δ in $\mathcal{O}\left(\frac{\sqrt{V^{k,f,g}}}{\delta}\right)$ queries, with $V^{k,f,g} = \mathbb{V}^{\mu^G}(\phi_g^{f,k}(H))$.

As described in Section 3.1, the AIS estimator (10) leads to a precision of δ for $n = \mathcal{O}\left(\frac{v^{k,f,g}(\theta^*)}{\delta^2}\right)$ samples, which is asymptotically less efficient than the above quantum estimator. However, no quantum advantage has been evidenced on current hardware for general Monte Carlo estimations and further challenges with respect to the precision of the evaluation of the integrand $\phi_g^{f,k}$ are expected in our specific context. Keeping working on the optimization of the estimators in the classical set-up while keeping track of the progress made on the development of quantum hardware therefore appears a reasonable path to follow.

6. Conclusions and Further Work

By leveraging on the approach proposed in [13] for quantile estimation, we have introduced in this paper an AIS variance reduction method for the computation of group-based convolution operators, a key component of equivariant neural networks. We have in particular used information geometry concepts to define an efficient update rule to infer the optimal sampling parametric distribution and have also shown promising results when working with the two-dimensional rotation group $SO(2)$ and von Mises distributions.

Further work will include the study of non-compact groups such as $SU(1,1)$ as to improve the efficiency of the computations underlying to the ENN introduced in [9]. As shown in [31], Souriau Thermodynamics can be used to build Gaussian distributions over $SU(1,1)$, which appear as natural candidates for applying the AIS scheme presented in this paper.

We have also seen that Monte Carlo computations can generally benefit from a quadratic speed-up in a quantum computing set-up. Further work will include the study of using AIS in this context as to provide a generic and efficient quantum algorithm for group-convolution computation. Benchmarking with group-Fourier transform-based approaches such as [21], which are more theoretically involved but with a promise of exponential speed-up, will also be of high interest, as it will be the case for results coming from the emerging field of quantum geometric deep learning [32,33].

Author Contributions: All authors contributed equally to the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is the result of some research work conducted by the authors at Thales Group.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bronstein, M.M.; Bruna, J.; Cohen, T.; Veličković, P. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. *arXiv* **2021**, arXiv:2104.13478.
2. Finzi, M.; Stanton, S.; Izmailov, P.; Wilson, A.G. Generalizing Convolutional Neural Networks for Equivariance to Lie Groups on Arbitrary Continuous Data. *arXiv* **2020**, arXiv:2002.12880.
3. Lafarge, M.W.; Bekkers, E.J.; Pluim, J.P.W.; Duits, R.; Veta, M. Roto-Translation Equivariant Convolutional Networks: Application to Histopathology Image Analysis. *arXiv* **2020**, arXiv:2002.08725.
4. Cohen, T.S.; Geiger, M.; Köhler, J.; Welling, M. Spherical CNNs. *arXiv* **2018**, arXiv:1801.10130.
5. Gerken, J.E.; Aronsson, J.; Carlsson, O.; Linander, H.; Ohlsson, F.; Petersson, C.; Persson, D. Geometric Deep Learning and Equivariant Neural Networks. *arXiv* **2021**, arXiv:2105.13926.
6. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
7. Chen, S.; Dobriban, E.; Lee, J.H. A Group-Theoretic Framework for Data Augmentation. *arXiv* **2020**, arXiv:1907.10905.
8. Kondor, R.; Trivedi, S. On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; Dy, J., Krause, A., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2018; Volume 80, pp. 2747–2755.
9. Lagrave, P.Y.; Cabanes, Y.; Barbaresco, F. SU(1,1) Equivariant Neural Networks and Application to Robust Toeplitz Hermitian Positive Definite Matrix Classification. In *Geometric Science of Information*; Nielsen, F., Barbaresco, F., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 577–584.
10. Finzi, M.; Welling, M.; Wilson, A.G. A Practical Method for Constructing Equivariant Multilayer Perceptrons for Arbitrary Matrix Groups. *arXiv* **2021**, arXiv:2104.09459.
11. Cohen, T.S.; Geiger, M.; Weiler, M. A General Theory of Equivariant CNNs on Homogeneous Spaces. In *Advances in Neural Information Processing Systems*; Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: New York, NY, USA, 2019; Volume 32, pp. 9145–9156.
12. Cohen, T.; Welling, M. Group Equivariant Convolutional Networks. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; Balcan, M.F., Weinberger, K.Q., Eds.; PMLR: New York, NY, USA, 2016; Volume 48, pp. 2990–2999.
13. Egloff, D.; Leippold, M. Quantile estimation with adaptive importance sampling. *Ann. Stat.* **2010**, *38*, 1244–1278. [[CrossRef](#)]
14. Ollivier, Y.; Arnold, L.; Auger, A.; Hansen, N. Information-Geometric Optimization Algorithms: A Unifying Picture via Invariance Principles. *J. Mach. Learn. Res.* **2011**, *18*, 1–65. [[CrossRef](#)]
15. Lagrave, P.Y.; Barbaresco, F. Generalized SU(1,1) Equivariant Convolution on Fock-Bargmann Spaces for Robust Radar Doppler Signal Classification. 2021, *working paper or preprint*.
16. von Mises, R. Über die “Ganzzahligkeit” der Atomgewicht und verwandte Fragen. *Phys. Z.* **1981**, *19*, 490–500.
17. Amari, S.I.; Barndorff-Nielsen, O.E.; Kass, R.E.; Lauritzen, S.L.; Rao, C.R. *Differential Geometry in Statistical Inference*; Lecture Notes–Monograph Series; Institute of Mathematical Statistics: Hayward, CA, USA, 1987; Volume 10, pp. 1–240.
18. Montanaro, A. Quantum speedup of Monte Carlo methods. *Proc. R. Soc. Math. Phys. Eng. Sci.* **2015**, *471*, 20150301. [[CrossRef](#)] [[PubMed](#)]
19. Herbert, S. Quantum Monte-Carlo Integration: The Full Advantage in Minimal Circuit Depth. *arXiv* **2021**, arXiv:2105.09100.
20. An, D.; Linden, N.; Liu, J.P.; Montanaro, A.; Shao, C.; Wang, J. Quantum-accelerated multilevel Monte Carlo methods for stochastic differential equations in mathematical finance. *Quantum* **2021**, *5*, 481. [[CrossRef](#)]
21. Castelazo, G.; Nguyen, Q.T.; Palma, G.D.; Englund, D.; Lloyd, S.; Kiani, B.T. Quantum algorithms for group convolution, cross-correlation, and equivariant transformations. *arXiv* **2021**, arXiv:2109.11330.
22. Botev, Z.; Ridder, A. Variance Reduction. In *Wiley StatsRef: Statistics Reference Online*; Wiley: Hoboken, NJ, USA, 2017; pp. 1–6. [[CrossRef](#)]
23. Jourdain, B. Adaptive variance reduction techniques in finance. *Adv. Financ. Model.* **2009**, *8*, 205.
24. Robbins, H.; Monroe, S. A Stochastic Approximation Method. *Ann. Math. Stat.* **1951**, *22*, 400–407. [[CrossRef](#)]
25. Martin, S.; Lagrave, P.Y. On the Benefits of SO(3)-Equivariant Neural Networks for Spherical Image Processing. 2022, *working paper or preprint*.
26. Kent, J.T. The Fisher-Bingham Distribution on the Sphere. *J. R. Stat. Soc. Ser. (Methodol.)* **1982**, *44*, 71–80. [[CrossRef](#)]
27. Niederreiter, H. *Random Number Generation and Quasi-Monte Carlo Methods*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1992. [[CrossRef](#)]
28. Sobol’, I. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput. Math. Math. Phys.* **1967**, *7*, 86–112. [[CrossRef](#)]
29. Orús, R.; Mugel, S.; Lizaso, E. Quantum computing for finance: Overview and prospects. *Rev. Phys.* **2019**, *4*, 100028. [[CrossRef](#)]
30. Brassard, G.; Høyer, P.; Mosca, M.; Tapp, A. Quantum amplitude amplification and estimation. *Contemp. Math.* **2002**, *305*, 53–74. [[CrossRef](#)]
31. Barbaresco, F. Lie Group Statistics and Lie Group Machine Learning Based on Souriau Lie Groups Thermodynamics & Koszul-Souriau-Fisher Metric: New Entropy Definition as Generalized Casimir Invariant Function in Coadjoint Representation. *Entropy* **2020**, *22*, 642. [[CrossRef](#)] [[PubMed](#)]

-
32. Larocca, M.; Sauvage, F.; Sbahi, F.M.; Verdon, G.; Coles, P.J.; Cerezo, M. Group-Invariant Quantum Machine Learning. *arXiv* **2022**, arXiv:2205.02261.
 33. Meyer, J.J.; Mularski, M.; Gil-Fuster, E.; Mele, A.A.; Arzani, F.; Wilms, A.; Eisert, J. Exploiting symmetry in variational quantum machine learning. *arXiv* **2022**, arXiv:2205.06217.