Kernel Design for Gaussian Mixture Models in Direct Policy Search

Lukas P. Fröhlich
ETH Zurich
lukasfro@ethz.ch

Leonel Rozo
Bosch Center for AI
leonel.rozo@.de.bosch.com

Melanie N. Zeilinger
ETH Zurich
mzeilinger@ethz.ch

***Abstract*—In this paper, we consider the problem of direct policy search with policies represented by Gaussian mixture models (GMMs). Due to its superior sample-efficiency, Bayesian optimization (BO) has become a prominent to address robot learning tasks. However, ignoring the respective geometry of the policy parameter space can impair the sample-efficiency. Hence, we propose to use a metric that acknowledges the space of probability distributions. In particular, we construct a novel stationary kernel based on the so-called probability product kernel that can be used in Gaussian processes – a key ingredient for BO. In preliminary experiments, we demonstrate that our approach is able to address the aforementioned issues and exhibits high potential for further research.**

I. Introduction

Sample-efficiency is one of the key criteria for robot learning in the real world to avoid wear and tear on the robot as well as to reduce time for the required oversight during the learning process. One particularly sample-efficient approach to the direct policy search problem is BO [2], in which the policy parameters are directly optimized based on the performance from rollouts on the real system. A key component of BO is a probabilistic surrogate model (e.g., Gaussian process (GP)) approximating the unknown objective [3]. In the context of BO, the GP is typically based on stationary kernels, i.e., kernels that only depend on the distance between its arguments $k(\boldsymbol{\theta}, \boldsymbol{\theta}') = k(d(\boldsymbol{\theta}, \boldsymbol{\theta}'))$, where $d(\cdot, \cdot)$ denotes a metric on the parameter space $\boldsymbol{\Theta} \ni \boldsymbol{\theta}, \boldsymbol{\theta}'$. The particular choice of d should reflect the underlying geometry of $\boldsymbol{\Theta}$.

Contributions: In this paper, we develop a metric tailored to GMMs with the purpose of improving BO's performance for direct policy search. In particular, the proposed metric can be computed efficiently and in closed-form, is numerically more stable compared to measures based on the KL-divergence, readily identifies the most relevant mixture component of the policy and implicitly reduces the search space's volume by exploiting the underlying geometry of the parameter space.

Related Work: Several works have considered non-Euclidean metrics in kernel methods: Moreno et al. [7] propose to use Jeffrey's divergence between GMMs, but already mention the numerical instabilities of the approach. Wilson et al. [9] compute the distance between policy parameters in terms of a KL divergence-based metric w.r.t. the resulting state distributions. More recently, Jaquier et al. [5] extended BO to Riemannian manifolds such that e.g. the geometry of positive definite matrices is properly accounted for.

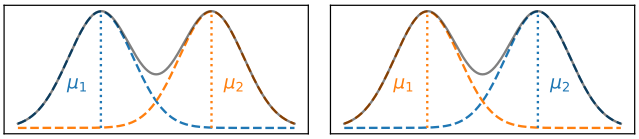

Fig. 1: Exemplary 1-dim. Gaussian Mixture Model.

Preliminaries: The parameter vector $\boldsymbol{\theta}$ representing the policy contains the respective mixture components' prior probabilities, mean vectors and covariance matrices of a GMM. By conditioning on a subset of the random variables, one readily obtains a predictive model to encode the robot's behaviour [1]. Direct policy search seeks parameters $\boldsymbol{\theta}^*$ that optimize a particular performance measure $f(\boldsymbol{\theta})$, to which end we employ BO. For a tutorial on BO, we refer the reader to [3].

II. Kernel Design for Gaussian Mixtures

Consider the following simple example in Fig. 1, for which the Euclidean metric fails to measure an appropriate distance between GMMs: given are two 1-dimensional GMMs, each with two mixture components of equal prior probability and variance, but different means such that $\boldsymbol{\theta} = [\mu_1, \mu_2]$ (left) and $\boldsymbol{\theta}' = [\mu_2, \mu_1]$ (right), respectively. In fact, both models represent the same underlying distribution, but the Euclidean distance is non-zero. However, not every appropriate metric (or divergence) between distributions can be used to construct a valid kernel that fulfills Mercer's condition [8]. Generally, it is difficult to judge a kernel's validity, but it is straightforward to build complex kernels from simpler ones via the kernel construction rules. Arguably the simplest kernel for inner product spaces is just the inner product itself $\langle \boldsymbol{\theta}_i, \boldsymbol{\theta}_j \rangle_{\boldsymbol{\Theta}}$. Hence, a necessary (but not sufficient) condition for a metric leading to a valid stationary kernel is that it is norm-induced, i.e., $d(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sqrt{\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_{\boldsymbol{\Theta}}}$, where $\|\cdot\|_{\boldsymbol{\Theta}}$ is the norm induced by the inner product $\langle \cdot, \cdot \rangle_{\boldsymbol{\Theta}}$.

In this paper, we are concerned with the distance between probability distributions, GMMs in particular. To this end, we construct a novel kernel based on the so-called *probability product kernel (PPK)* [6]

$$k_{\mathrm{PPK}}(p_{\boldsymbol{\theta}}, p_{\boldsymbol{\theta}'}) = \int p_{\boldsymbol{\theta}}(\mathbf{x})^{\rho} p_{\boldsymbol{\theta}'}(\mathbf{x})^{\rho} \, \mathrm{d}\mathbf{x} = \langle p_{\boldsymbol{\theta}}^{\rho}, p_{\boldsymbol{\theta}'}^{\rho} \rangle_{L^2}, \quad (1)$$

for $\rho \in \mathbb{R}_{\geq 0}$, which, in fact, is the definition of an inner product between $p_{\boldsymbol{\theta}}^{\rho}, p_{\boldsymbol{\theta}'}^{\rho}$ in the space of square-integrable functions L^2.

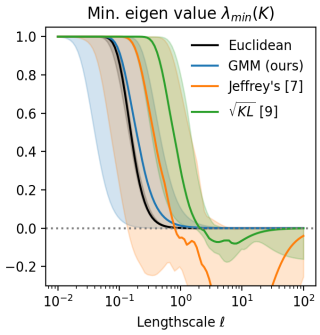

Fig. 2: Min. EV of the Gram matrix for different distances.

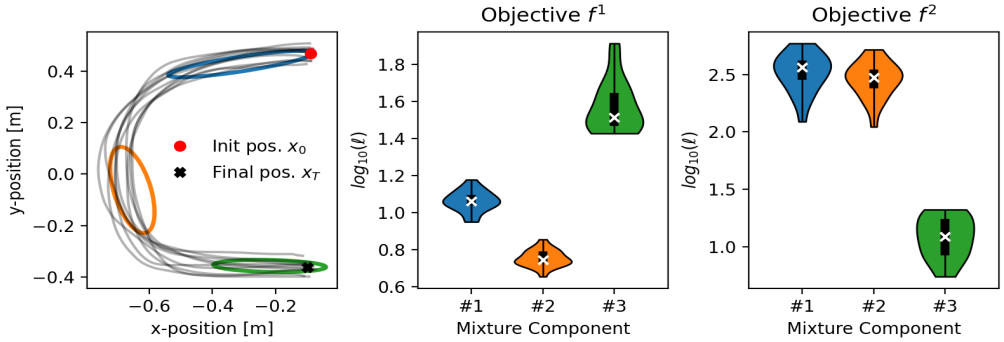

Fig. 3: Left: Visualization of the (marginal) GMM and sampled resulting end-effector trajectories. Center/Right: Inferred lengthscales for f^1 and f^2 from 20 independent experiments, respectively.

For the remainder of this paper, we assume $\rho = 1$ and for ease of notation $k(p_{\boldsymbol{\theta}}, p_{\boldsymbol{\theta}'}) = k(\boldsymbol{\theta}, \boldsymbol{\theta}')$. For GMMs, we can compute (1) analytically such that

$$d_{\text{GMM}}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sqrt{\langle \boldsymbol{\theta}, \boldsymbol{\theta} \rangle + \langle \boldsymbol{\theta}', \boldsymbol{\theta}' \rangle - 2\langle \boldsymbol{\theta}, \boldsymbol{\theta}' \rangle}, \text{ with} \quad (2)$$

$$\langle \boldsymbol{\theta}, \boldsymbol{\theta}' \rangle = \sum_{k=1}^{K} \sum_{k'=1}^{K'} \pi_k \pi_{k'} \mathcal{N}(\boldsymbol{\mu}_k | \boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_k + \boldsymbol{\Sigma}_{k'}). \quad (3)$$

With the definition above, each mixture component's influence on the distance is mostly determined by its covariance matrix, i.e., low-variance components contribute more to the distance and vice versa for components with high variance. However, the GP's dependent variable might be highly sensitive to small changes in the mean for high-variance components. We therefore introduce lengthscale parameters in our metric to account for this shortcoming of (3) and as such enable automatic relevance determination (ARD) for the GP model,

$$\langle \boldsymbol{\theta}, \boldsymbol{\theta}' \rangle^{\Lambda} = \sum_{k=1}^{K} \sum_{k'=1}^{K} \pi_k \pi_{k'} \mathcal{N}\Big(\frac{\boldsymbol{\mu}_k}{\ell_k} | \frac{\boldsymbol{\mu}'_k}{\ell'_k}, \boldsymbol{\Sigma}_k + \boldsymbol{\Sigma}_{k'} \Big), \quad (4)$$

with $\Lambda = [\ell_1, \dots, \ell_K]$ denoting the lengthscales (inverse relevance parameters) for each of the K mixture components.

III. Preliminary Results

Numerical Stability: We demonstrate that the proposed metric (2) leads to a proper kernel resulting in a numerically stable GP model as opposed to other metrics based on the KL-divergence [7, 9]. To this end, we sample 20 3-dim. GMMs and compute the minimum eigenvalue for the respective Gram matrices $\mathbf{K}$ using the RBF kernel but different metrics and varying lengthscale parameters. Since the KL divergence between GMMs has no analytical form, we use standard approximation techniques [4]. The results are shown in Fig. 2, where the solid line and shaded areas represent the median and 10^{th}/90^{th} percentiles across 20 independent experiments, respectively. Note that the respective minimum eigenvalues of the Gram matrices for the KL-based distances can become negative, hence violating the positive definiteness of kernels. In contrast, our approach leads to a valid kernel.

Automatic Relevance Determination: We demonstrate that the introduced lengthscale parameters in (4) are able to identify the most relevant mixture components for a given task. In particular, we simulate rollouts for a planar 2-DoF robotic arm following random policies (see gray lines in Fig. 3 (left), indicating end-effector position) and evaluate each trajectory by two different performance measures: $f^1(\boldsymbol{\theta}) = \min_{t \in [0,T]} \det \mathbf{J}_t^\top \mathbf{J}_t$ with $\mathbf{J}_t$ being the Jacobian of the robot's joint configuration at time t, $f^2(\boldsymbol{\theta}) = \|\mathbf{x}_T - \hat{\mathbf{x}}\|_2$ measuring the end-effector's position $\mathbf{x}_T$ at the end of the rollout with respect to a reference point $\hat{\mathbf{x}}$. Note that for the respective objectives, different components of the underlying GMM have larger influence. For f^1, the position of the second (orange) component gives rise to joint configurations close to singularities whereas for f^2 the last (green) component determines the performance. Based on the sampled policies $\boldsymbol{\theta}_i$ and corresponding function values $f_i^{1,2}$, we fit a GP with RBF kernel using our extended metric based on (2) and (4) by maximizing the marginal log-likelihood. The respective lengthscale parameters for both objectives from 20 independent experiments are shown in Fig. 3 (center and right). Our model reliably identifies the most relevant components.

IV. Conclusion and Outlook

In this work, we propose a novel kernel that considers the geometry of the space of GMMs. The application of this method targets direct policy search with BO, in which the GP can leverage the new kernel. We present promising results in terms of numerical stability and detection of most relevant mixture components. Future research focuses on deploying this approach for actual policy search problems.

References

[1] S. Calinon. A tutorial on task-parameterized movement learning and retrieval. *Intelligent Service Robotics*, 9(1):1–29, 2016.

[2] K. Chatzilygeroudis, V. Vassiliades, F. Stulp, S. Calinon, and J.-B. Mouret. A survey on policy search algorithms for learning robot controllers in a handful of trials. *IEEE Trans. Robot.*, 2019.

[3] P.I. Frazier. A tutorial on Bayesian optimization. *arXiv:1807.02811*, 2018.

[4] J.R. Hershey and P.A. Olsen. Approximating the KL divergence between gaussian mixture models. In *ICASSP*, 2007.

[5] N. Jaquier, L. Rozo, S. Calinon, and M. Buerger. Bayesian optimization meets Riemannian manifolds in robot learning. In *CORL*, 2019.

[6] T. Jebara, R. Kondor, and A. Howard. Probability Product Kernels. *JMLR*, 5:819–844, 2004.

[7] P.J. Moreno, P.P. Ho, and N. Vasconcelos. A KL divergence based kernel for SVM classification in multimedia applications. In *NeurIPS*, 2004.

[8] J. Shawe-Taylor and N. Cristianini. *Kernel methods for pattern analysis*. Cambridge university press, 2004.

[9] A. Wilson, A. Fern, and P. Tadepalli. Using trajectory data to improve Bayesian optimization for reinforcement learning. *JMLR*, 15(1):253–282, 2014.