A Experiments

A.1 Experimental Details

All the robotic experiments reported in Section 4.2 used 5 initial random samples before starting the Bayesian optimization loop. For the orientation experiment, the parameters of the cost $f(\boldsymbol{q})$ were fixed to $\tilde{\boldsymbol{q}} = (0.8 \quad 0.6 \quad 0 \quad 0)^\mathsf{T}$, $w_{\boldsymbol{q}} = 1.0$, $w_{\boldsymbol{\tau}} = 1e^{-4}$, and $w_{\mathbf{M}} = 1$. We used a velocity-based orientation controller $\dot{\boldsymbol{\theta}}_r = \mathbf{J}(\boldsymbol{\theta})^\dagger \left[\dot{\boldsymbol{q}}_d + \mathbf{K}_p \operatorname{Log}_{\tilde{\boldsymbol{q}}}(\boldsymbol{q})\right]$, with $\boldsymbol{\theta}$ being the robot joint position, $\mathbf{J}(\boldsymbol{\theta})$ representing the robot Jacobian, and $\dagger$ denoting the pseudo-inverse operator. Furthermore, $\operatorname{Log}(\cdot)$ measures the quaternion tracking error on the sphere manifold, and $\mathbf{K}_p = 6.0\mathbf{I}_4$.

For the manipulability optimization task, the weights in the cost $f(\widehat{\mathbf{M}})$ were set to $w_{\ddot{\boldsymbol{p}}} = 0.1$, $w_{\mathbf{M}} = 0.1$, and $w_{\boldsymbol{t}} = 15$. We used velocity-based control of the form $\dot{\boldsymbol{\theta}}_r = \mathbf{J}(\boldsymbol{\theta})^\dagger \left[\dot{\boldsymbol{x}}_d + \mathbf{K}_p(\boldsymbol{x}_d - \boldsymbol{x})\right] + \mathbf{N}(\boldsymbol{\theta})\mathcal{J}(\mathbf{M})^\dagger_{(3)} \left[\mathbf{K}_M \operatorname{vec}\left(\operatorname{Log}_{\widehat{\mathbf{M}}}(\mathbf{M})\right)\right]$, with $\boldsymbol{\theta}$ and $\boldsymbol{x}$ denoting the robot joint and Cartesian position, while $\mathbf{J}(\boldsymbol{\theta})$ is the robot Jacobian and $\mathbf{N}(\boldsymbol{\theta})$ is the Jacobian nullspace. Moreover, $\operatorname{Log}(\cdot)$ measures the manipulability tracking error on the SPD manifold, which is mapped to joint velocities via the pseudo-inverse of the manipulability Jacobian $\mathcal{J}(\mathbf{M})$. In addition, $\operatorname{vec}(\cdot)$ denotes the vectorization of a matrix, and $(\cdot)_{(3)}$ denotes the 3-mode matricization of a tensor. We refer the interested reader to Jaquier et al. [26] for a complete description of the manipulability controller. The position and manipulability tracking gains, for the main and secondary control tasks, correspond to $\mathbf{K}_p = 75\mathbf{I}_2$ and $\mathbf{K}_M = 7.0\mathbf{I}_3$.

Finally, for the simple path planning problem, we considered a 9×9 grid as a discretization of the robot planar environment, which included a circle-shaped obstacle of radius 1 centered in $(4, 4)$. The nodes inter-distance was 1 unit. The simple graph to optimize consisted of a single branch of 5 nodes connecting the robot start location to the goal. The parameters of the cost $f(\boldsymbol{x}_{1,\dots,m})$ were fixed to $w_g = 1$, and $w_d = 25$.

A.2 Additional Benchmark Experiments

Figure 4 shows the performance results for additional benchmark functions projected onto the manifolds $\mathbb{S}^3$, $\mathbb{S}^5$, $\mathrm{SO}(3)$, $\mathcal{S}^2_{++}$, and $\mathcal{H}^3$. The left graphs show the evolution of the median for the different Bayesian optimization approaches, and the right graphs display the distribution of the logarithm of the simple regret of the final recommendation $\boldsymbol{x}_N$ after 200 iterations. Similarly to the experiments presented in the main paper, we observe that geometry-aware algorithms generally match or outperform their Euclidean counterparts, resulting in better solution quality, faster convergence rate, and lower variance. These differences are more pronounced in higher dimensions, for instance when comparing $\mathbb{S}^5$ against to $\mathbb{S}^3$, and for manifolds which differ significantly from the Euclidean space.

A.3 Additional Robotics Experiments

Figure 5 shows the performance results for two additional robotics experiments. Similarly to the first experiment of Section 4.2, we first use Bayesian optimization as an orientation sampler, where a velocity-controlled robot samples an orientation reference $\boldsymbol{x} = \hat{\boldsymbol{q}}$ around a prior orientation $\tilde{\boldsymbol{q}}$. The objective of the robot is to minimize the error between the prior and the current end-effector orientation with low joint torques and an isotropic manipulability ellipsoid, i.e., $f(\boldsymbol{q}) = \sum w_{\boldsymbol{q}} \operatorname{dist}_{\mathcal{S}^3}(\tilde{\boldsymbol{q}}, \boldsymbol{q}) + w_{\boldsymbol{\tau}} \|\boldsymbol{\tau}\|_2^2 + w_{\mathbf{M}} \operatorname{cond}(\mathbf{M})$, where $\boldsymbol{q}$ is the current end-effector orientation, $\boldsymbol{\tau}$ is the joint torques, and $\operatorname{cond}(\mathbf{M})$ is the condition number of the velocity manipulability ellipsoid. The parameters of the cost $f(\boldsymbol{q})$ were fixed to $\tilde{\boldsymbol{q}} = (0.8 \quad 0.6 \quad 0 \quad 0)^\mathsf{T}$, $w_{\boldsymbol{q}} = 1.0$, $w_{\boldsymbol{\tau}} = 0.1$, and $w_{\mathbf{M}} = 1e^{-4}$.

Next, we consider a task-compatible manipulability optimization scenario, similar to the second experiment of Section 4.2. Namely, a 8-degree-of-freedom planar robot is required to track a desired Cartesian velocity trajectory leading to a vertical line, while tracking a desired manipulability ellipsoid in its nullspace. The optimizer aims at finding the desired manipulability $\boldsymbol{x} = \widehat{\mathbf{M}} \in \mathcal{S}^2_{++}$ so that the end-effector acceleration $\ddot{\boldsymbol{p}}$ is minimized and the robot tracks the desired manipulabity with a high

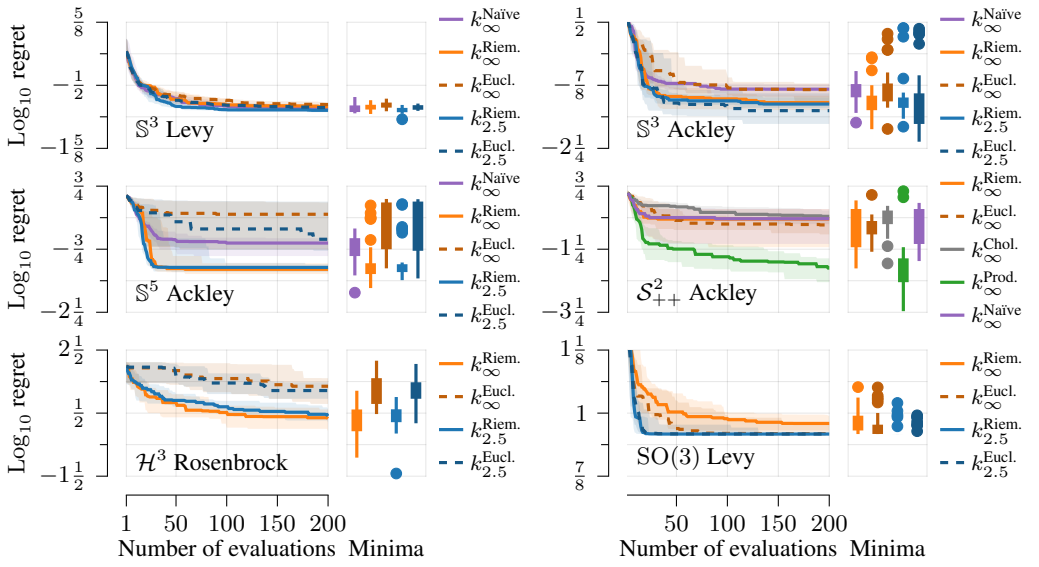

Figure 4: Logarithm of the regret for a set of additional benchmark test functions on the manifolds $\mathbb{S}^3$, $\mathbb{S}^5$, $\mathrm{SO}(3)$, $\mathcal{S}^2_{++}$, and $\mathcal{H}^3$, for both geometry-aware and Euclidean Bayesian optimization algorithms.

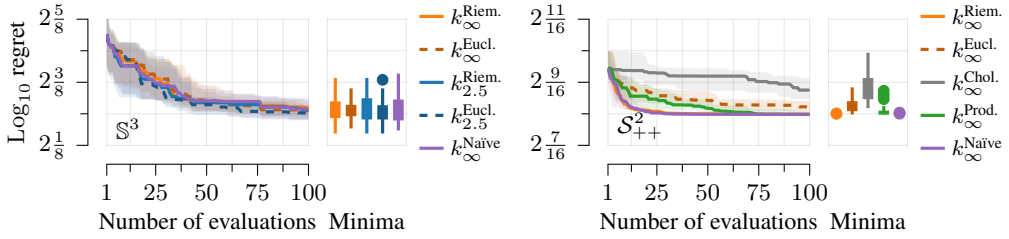

Figure 5: Logarithm of the regret for the additional orientation, and task-compatible manipulability problems, for both geometry-aware and Euclidean Bayesian optimization algorithms.

precision. This leads to the cost function $f(\widehat{\mathbf{M}}) = \sum w_{\ddot{\boldsymbol{p}}} \|\ddot{\boldsymbol{p}}\|_2 + w_{\mathbf{M}} \operatorname{dist}_{\mathcal{S}^2_{++}}(\mathbf{M}, \widehat{\mathbf{M}})$ with $\mathbf{M}$ the current manipulability. Here, the weights are given the following values $w_{\ddot{\boldsymbol{p}}} = 0.1$ and $w_{\mathbf{M}} = 1.0$.

As shown in Figure 5, geometry-aware Bayesian optimization algorithms match or outperform their Euclidean counterpart, mirroring the results presented in the main paper. The performance differences are more pronounced in the experiment presented in Section 4.2 compared to this second orientation experiment due to the uneven landscape of the cost function of the former experiment. The results of the manipulability experiment are similar to those presented in Section 4.2. The naïve Riemannian squared exponential kernel—which for certain length scales may become ill-defined—performs competitively in this setting. This may occur because of the low dimensionality of the problem, and because small length scales keep it out of problematic situations.

B Theory

B.1 Matérn kernels on general Riemannian manifolds and their connection to heat kernels

The theoretical framework defining appropriate notions of stochastic partial differential equations (SPDEs) and Matérn kernels as kernels of the unique solutions of these equations was presented in Appendix D of Borovitskiy et al. [8] for the case of compact Riemannian manifolds. In the non-compact case, the technique that yields explicit solutions of the SPDEs as series depending on Laplace–Beltrami eigenpairs does not apply, but the implicit definition of Matérn Gaussian processes and the corresponding kernel remain valid under mild regularity conditions.

Specifically, if we restrict our attention to connected complete Riemannian manifolds of bounded geometry, which is indeed a rather nonrestrictive assumption, then one can generalize both the functional calculus formalism and the definitions of the appropriate function spaces needed to make the SPDEs rigorously defined. This can be done by means of general spectral theory, which is reviewed by De Vito et al. [12]. The theory of Borovitskiy et al. [8] in this general setting shows that the Riemannian Matérn and squared exponential kernels are the reproducing kernels of Sobolev and diffusion spaces studied by De Vito et al. [12]. Because of this, Theorem 8 from the latter work directly implies that the Riemannian squared exponential kernel as defined by (the generalized theory of) Borovitskiy et al. [8] coincides with the heat kernels as fundamental solutions of the heat (diffusion) equation on the Riemannian manifold. This is a crucial point for this work, since it allows us to connect to the existing literature on heat kernels in some interesting non-compact cases.

B.2 Matérn kernels as integrals of heat kernels

By Appendix B.1, Riemannian squared exponential and Matérn kernels are well defined on sufficiently regular non-compact manifolds, and Riemannian squared exponential kernels coincide with heat kernels, which are widely studied in the mathematics literature even in non-compact cases. Because of this, it is natural to seek a way to express Riemannian Matérn kernels in terms of the corresponding heat kernels.

In the Euclidean case, this is given by equation (2), which we prove in Appendix B.2.1. A similar but slightly different relation holds in the compact Riemannian case, which we prove in Appendix B.2.2. For the general Riemannian case we adopt (2) as the *definition* of Matérn kernels. We prove that the kernels defined this way are positive definite in Appendix B.2.3.

B.2.1 Euclidean Matérn kernels as integral of heat kernels

The Euclidean Matérn and squared exponential kernels are given by the following formulas [39]:

$$k_{\nu,\kappa,\sigma^2}(\boldsymbol{x},\boldsymbol{x}') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \Big(\sqrt{2\nu}\rho/\kappa\Big)^{\nu} K_\nu\Big(\sqrt{2\nu}\rho/\kappa\Big), \qquad k_{\infty,\kappa,\sigma^2}(\boldsymbol{x},\boldsymbol{x}') = \sigma^2 e^{-\frac{\rho^2}{2\kappa^2}}, \tag{10}$$

where $\rho = \|\boldsymbol{x} - \boldsymbol{x}'\|$ denotes the distance between a pair of inputs. These kernels may be represented as inverse Fourier transforms of the corresponding spectral densities S_{ν,κ,σ^2} [39], that is

$$k_{\nu,\kappa,\sigma^2}(\boldsymbol{x},\boldsymbol{x}') = \int_{\mathbb{R}^d} S_{\nu,\kappa,\sigma^2}(\|\boldsymbol{\xi}\|) e^{2\pi i \langle \boldsymbol{x}-\boldsymbol{x}', \boldsymbol{\xi}\rangle} \,\mathrm{d}\boldsymbol{\xi}. \tag{11}$$

This covers both cases, whether ν is finite or infinite. The respective spectral densities are given by

$$S_{\nu,\kappa,\sigma^2}(\lambda) = \sigma^2 \overbrace{\frac{2^d \pi^{d/2} \Gamma(\nu + d/2)(2\nu)^\nu}{\Gamma(\nu)\kappa^{2\nu}}}^{\text{denote by } 1/C_\nu} \left(\frac{2\nu}{\kappa^2} + 4\pi^2\lambda^2\right)^{-\nu-\frac{d}{2}}, \tag{12}$$

$$S_{\infty,\kappa,\sigma^2}(\lambda) = \sigma^2 \underbrace{(2\pi\kappa^2)^{d/2}}_{\text{denote by } 1/C_\infty} e^{-2\pi^2\kappa^2\lambda^2}. \tag{13}$$

Here d is the dimension of the Euclidean space under consideration and C_ν, C_∞ are the normalizing constants that ensure $k_{\nu,\kappa,\sigma^2}(\boldsymbol{x},\boldsymbol{x}) = \sigma^2$. Note that, despite this notation, they depend on both ν and κ.

We start verifying equation (2) by noting that by Gradshteyn and Ryzhik [18], Section 3.326, Item 2 we have

$$\int_0^\infty u^n e^{-au} \,\mathrm{d}u = \Gamma(n+1) a^{-n-1}. \tag{14}$$

Substituting $n = \nu + d/2 - 1$ and $a = 2\nu/\kappa^2 + 4\pi^2\lambda^2$ into this equation and then performing a simple rearrangement of terms, we get the following expression for $C_\nu/\sigma^2 S_{\nu,\kappa,\sigma^2}(\lambda)$:

$$\left(\frac{2\nu}{\kappa^2} + 4\pi^2\lambda^2\right)^{-\nu-\frac{d}{2}} = \Gamma(\nu + d/2)^{-1} \int_0^\infty u^{\nu+\frac{d}{2}-1} e^{-\frac{2\nu}{\kappa^2}u} e^{-4\pi^2\lambda^2 u}\,\mathrm{d}u, \tag{15}$$

$$= (4\pi)^{-\frac{d}{2}}\Gamma(\nu + d/2)^{-1} \int_0^\infty u^{\nu-1} e^{-\frac{2\nu}{\kappa^2}u} (4\pi u)^{\frac{d}{2}} e^{-4\pi^2\lambda^2 u}\,\mathrm{d}u, \tag{16}$$

$$= \sigma^{-2}(4\pi)^{-\frac{d}{2}}\Gamma(\nu + d/2)^{-1} \int_0^\infty u^{\nu-1} e^{-\frac{2\nu}{\kappa^2}u} S_{\infty,\sqrt{2u},\sigma^2}(\lambda)\,\mathrm{d}u. \tag{17}$$

Now, using equations (11) and (12) and then (17), we write

$$k_{\nu,\kappa,\sigma^2}(\boldsymbol{x}, \boldsymbol{x}') = \frac{\sigma^2}{C_\nu} \int_{\mathbb{R}^d} \left(\frac{2\nu}{\kappa^2} + 4\pi^2\|\boldsymbol{\xi}\|^2\right)^{-\nu-\frac{d}{2}} e^{2\pi i\langle \boldsymbol{x}-\boldsymbol{x}', \boldsymbol{\xi}\rangle}\,\mathrm{d}\boldsymbol{\xi} \tag{18}$$

$$= \frac{1}{C_\nu(4\pi)^{\frac{d}{2}}\Gamma(\nu + d/2)} \int_{\mathbb{R}^d} \int_0^\infty u^{\nu-1} e^{-\frac{2\nu}{\kappa^2}u} S_{\infty,\sqrt{2u},\sigma^2}(\|\boldsymbol{\xi}\|)\,\mathrm{d}u\, e^{2\pi i\langle \boldsymbol{x}-\boldsymbol{x}', \boldsymbol{\xi}\rangle} d\boldsymbol{\xi} = \ldots \tag{19}$$

By changing the order of integration, rearranging terms and using formulas (11) and (13) we get

$$\ldots = \frac{1}{C_\nu(4\pi)^{\frac{d}{2}}\Gamma(\nu + d/2)} \int_0^\infty u^{\nu-1} e^{-\frac{2\nu}{\kappa^2}u} k_{\infty,\sqrt{2u},\sigma^2}(\boldsymbol{x}, \boldsymbol{x}')\,\mathrm{d}u, \tag{20}$$

which coincides with (2) up to a multiplicative constant which we disregard for convenience since it only affects the normalization of the kernel.

B.2.2 Compact Riemannian Matérn kernels as integral of heat kernels

Normalizing constants in the definition of the Riemannian Matérn and squared exponential kernels are not available in closed form and, at the same time, depend on the length scale. To get around this, we connect the unnormalized versions of compact Riemannian Matérn kernels to their respective unnormalized squared exponential kernels. As a result, the relation turns out to be slightly different, but still very similar to (2).

First, we introduce the notation similar to (12) and (13), where the superscript u stands for *unnormalized*:

$$S^u_{\nu,\kappa}(\lambda) = \left(\frac{2\nu}{\kappa^2} + \lambda\right)^{-\nu-\frac{d}{2}} \qquad S^u_{\infty,\kappa}(\lambda) = e^{-\frac{\kappa^2}{2}\lambda}. \tag{21}$$

Then for both finite and infinite ν we have $k_{\nu,\kappa,\sigma^2}(x, x') = \sigma^2/C_\nu \sum_{n=0}^\infty S^u_{\nu,\kappa}(\lambda_n) f_n(x) f_n(x')$. We are going to relate the unnormalized kernels given, both for finite and infinite ν, by

$$k^u_{\nu,\kappa}(x, x') = \sum_{n=0}^\infty S^u_{\nu,\kappa}(\lambda_n) f_n(x) f_n(x'). \tag{22}$$

As in the Euclidean case we use (14), but this time with $n = \nu + d/2 - 1$ and $a = 2\nu/\kappa^2 + \lambda$, where n stays as before and a is slightly different, and then perform calculations similar to (17), obtaining

$$S^u_{\nu,\kappa}(\lambda) = \left(\frac{2\nu}{\kappa^2} + \lambda\right)^{-\nu-\frac{d}{2}} = \Gamma(\nu + d/2)^{-1} \int_0^\infty u^{\nu+\frac{d}{2}-1} e^{-\frac{2\nu}{\kappa^2}u} e^{-\lambda u}\,\mathrm{d}u \tag{23}$$

$$= \Gamma(\nu + d/2)^{-1} \int_0^\infty u^{\nu+\frac{d}{2}-1} e^{-\frac{2\nu}{\kappa^2}u} S^u_{\infty,\sqrt{2u}}(\lambda)\,\mathrm{d}u. \tag{24}$$

Now, write for finite ν

$$k^u_{\nu,\kappa}(x,x') = \sum_{n=0}^{\infty} S^u_{\nu,\kappa}(\lambda_n) f_n(x) f_n(x') \tag{25}$$

$$= \sum_{n=0}^{\infty} \Gamma(\nu + d/2)^{-1} \int_0^{\infty} u^{\nu+\frac{d}{2}-1} e^{-\frac{2\nu}{\kappa^2}u} S^u_{\infty,\sqrt{2u}}(\lambda)\,\mathrm{d}u f_n(x) f_n(x') \tag{26}$$

$$= \Gamma(\nu + d/2)^{-1} \int_0^{\infty} u^{\nu+\frac{d}{2}-1} e^{-\frac{2\nu}{\kappa^2}u} \sum_{n=0}^{\infty} S^u_{\infty,\sqrt{2u}}(\lambda) f_n(x) f_n(x')\,\mathrm{d}u \tag{27}$$

$$= \Gamma(\nu + d/2)^{-1} \int_0^{\infty} u^{\nu+\frac{d}{2}-1} e^{-\frac{2\nu}{\kappa^2}u} k^u_{\infty,\sqrt{2u}}(x,x')\,\mathrm{d}u \tag{28}$$

which yields the expression

$$k^u_{\nu,\kappa}(x,x') = \Gamma(\nu + d/2)^{-1} \int_0^{\infty} u^{\nu+\frac{d}{2}-1} e^{-\frac{2\nu}{\kappa^2}u} k^u_{\infty,\sqrt{2u}}(x,x')\,\mathrm{d}u \tag{29}$$

which is of the form we were looking for.

B.2.3 Integral Matérn kernels are positive (semi)definite whenever heat kernels are

Assume k_{ν,κ,σ^2} is defined by (2) and that the corresponding heat kernel $k_{\infty,\kappa,\sigma^2}$ is positive definite.[3] Take some locations $x_1, \ldots, x_n = \boldsymbol{x}$ and consider the matrix $\mathbf{K}^{\nu}_{\boldsymbol{xx}}$ with elements $k_{\nu,\kappa,\sigma^2}(x_i, x_j)$. In order to prove that k_{ν,κ,σ^2} is positive definite we need to show that $\mathbf{K}^{\nu}_{\boldsymbol{xx}}$ is a positive definite matrix for an arbitrary choice of n and $\boldsymbol{x}$. This means that we need to show that $\boldsymbol{y}^\top \mathbf{K}^{\nu}_{\boldsymbol{xx}} \boldsymbol{y}^\top > 0$ for all nonzero vectors $\boldsymbol{y} \in \mathbb{R}^n$.

To prove that $\mathbf{K}^{\nu}_{\boldsymbol{xx}}$ is positive definite, consider the matrices $\mathbf{K}^{\infty,\kappa}_{\boldsymbol{xx}}$ with elements $k_{\infty,\kappa,\sigma^2}(x_i, x_j)$. Then, extending equation (2) to matrices, we have

$$\mathbf{K}^{\nu}_{\boldsymbol{xx}} = \int_0^{\infty} u^{\nu-1} e^{-\frac{2\nu}{\kappa^2}u} K^{\infty,\sqrt{2u}}_{\boldsymbol{xx}}\,\mathrm{d}u. \tag{30}$$

Because of this, we obtain

$$\boldsymbol{y}^\top \mathbf{K}^{\nu}_{\boldsymbol{xx}} \boldsymbol{y} = \int_0^{\infty} \underbrace{u^{\nu-1} e^{-\frac{2\nu}{\kappa^2}u}}_{*} \underbrace{\boldsymbol{y}^\top K^{\infty,\sqrt{2u}}_{\boldsymbol{xx}} \boldsymbol{y}}_{**}\,\mathrm{d}u, \tag{31}$$

where the factor $*$ of the integrand is obviously positive and the factor $**$ of the integrand is positive because $\mathbf{K}^{\infty,\sqrt{2u}}_{\boldsymbol{xx}}$ is positive definite by assumption, thus the integral is positive. Thus k_{ν,κ,σ^2} is positive definite. The argument transfers to the case of positive semi-definiteness mutatis mutandis.

B.3 Matérn kernels for the torus

On the torus $\mathbb{T}^d$ (with flat metric), every vector of integers $\boldsymbol{\tau} \in \mathbb{Z}^d$ corresponds to the eigenvalue $\lambda_{\boldsymbol{\tau}} = 4\pi^2 \|\tau\|^2$ and to the two orthonormal eigenfunctions $f_{\boldsymbol{\tau},1}(\boldsymbol{x}) = \sqrt{2}\cos(2\pi\langle\boldsymbol{\tau}, \boldsymbol{x}\rangle)$ and $f_{\boldsymbol{\tau},2}(\boldsymbol{x}) = \sqrt{2}\sin(2\pi\langle\boldsymbol{\tau}, \boldsymbol{x}\rangle)$, unless of course $\boldsymbol{\tau} = 0$, which corresponds to the eigenvalue $\lambda_{\mathbf{0}} = 0$ and a single eigenfunction $f_{\boldsymbol{\tau}}(\boldsymbol{x}) = 1$: for details, see Gordon [17]. Note that the eigenvalues corresponding to distinct $\boldsymbol{\tau}, \boldsymbol{\tau}' \in \mathbb{Z}^d$ may coincide, take e.g. $\boldsymbol{\tau} = (15, 20)$ and $\boldsymbol{\tau'} = (7, 24)$. However, the eigenfunctions corresponding to distinct $\boldsymbol{\tau}, \boldsymbol{\tau}' \in \mathbb{Z}^d$ will always be different (and orthogonal) unless $\boldsymbol{\tau} = -\boldsymbol{\tau}$. Finally, observe that

$$f_{\boldsymbol{\tau},1}(\boldsymbol{x}) f_{\boldsymbol{\tau},1}(\boldsymbol{x}') + f_{\boldsymbol{\tau},2}(\boldsymbol{x}) f_{\boldsymbol{\tau},2}(\boldsymbol{x}') = 2\cos(2\pi\langle\boldsymbol{\tau}, \boldsymbol{x} - \boldsymbol{x}'\rangle) \tag{32}$$

for $\boldsymbol{\tau} \neq 0$ because of the identity $\cos(x - y) = \cos(x)\cos(y) + \sin(x)\sin(y)$.

Since the eigenpairs are parameterized by vectors of integers instead of just natural numbers, it is thus convenient to reorder the series in (1) accordingly—this can be done because the series converges

[3] Heat kernels on connected complete Riemannian manifolds of bounded geometry are positive definite, this follows, for example, from the discussion in Appendix B.1. Their approximations used for practical Gaussian process regression are typically positive semi-definite.

unconditionally [8]. Then, the summation should only be performed over *half* of $\mathbb{Z}^d$ to exclude the possibility of accounting non-orthogonal (even linearly dependent) eigenfunctions twice: for $\boldsymbol{\tau} \in \mathbb{Z}^d$, $\boldsymbol{\tau} \neq 0$ and for $-\boldsymbol{\tau}$. Instead of changing the summation index set, we notice that the expression on the right-hand side of (32) does not change when $\boldsymbol{\tau}$ is substituted for $-\boldsymbol{\tau}$ because cosine is an even function and divide the summands for $\boldsymbol{\tau} \neq 0$ by two, turning $2\cos(2\pi\langle\boldsymbol{\tau}, \boldsymbol{x} - \boldsymbol{x}'\rangle)$ into $\cos(2\pi\langle\boldsymbol{\tau}, \boldsymbol{x} - \boldsymbol{x}'\rangle)$. This leads to the expression of the form given by (3) for the Matérn and squared exponential kernels, namely

$$k_{\nu,\kappa,\sigma^2}(\boldsymbol{x}, \boldsymbol{x}') = \frac{\sigma^2}{C_\nu} \sum_{\boldsymbol{\tau} \in \mathbb{Z}^d} \Phi(4\pi^2 \|\tau\|^2) \cos(2\pi\langle\boldsymbol{\tau}, \boldsymbol{x} - \boldsymbol{x}'\rangle), \tag{33}$$

where Φ is as in (1). Note that the case of $\boldsymbol{\tau} = \mathbf{0}$ is properly accounted for because $\cos(2\pi\langle\mathbf{0}, \boldsymbol{x} - \boldsymbol{x}'\rangle) = 1$.

B.4 Heat and Matérn kernel for the special orthogonal group

The formula for the heat (squared exponential) kernel for the special orthogonal group $\mathrm{SO}(d)$ is available in the literature: see e.g. Wong [56], page 12, and Stein [47], page 45. As described in the main text, the expression for the kernel is

$$k_{\infty,\kappa,\sigma^2}(\mathbf{X}, \mathbf{Y}) = \frac{\sigma^2}{C_\infty} \sum_{\pi} e^{-\frac{\kappa^2}{2}\lambda_\pi} d_\pi \chi_\pi(\mathbf{X}\mathbf{Y}^{-1}). \tag{5}$$

We now discuss the precise meaning of the symbols π, λ_π, d_π, χ_π in equation (5) and how to compute their values in practice.

First, the summation variable π runs over the set of *highest weights*—these are effectively tuples of non-negative integers that enumerate *irreducible representations* of the group, i.e. homomorphisms $\pi\colon \mathrm{SO}(d) \to \mathrm{GL}(V)$ for some (complex) vector space V such that V cannot be split into a direct sum of non-trivial $\pi(\mathrm{SO}(d))$-invariant subspaces. Iterating over the highest weights amounts to iterating over the tuples of non-negative integers and then post-processing them according to the theory from Bröcker and Tom Dieck [9], Chapter VI, Section 2. To calculate the kernel to a given numerical precision, one can grade all such tuples by the sum of their entries, and only leave those terms of (5) which correspond to the entries with the smallest sum.

To each *highest weight* π there corresponds a particular eigenvalue λ_π of the Laplace–Beltrami operator. It can be computed as the value of a certain quadratic polynomial in the entries of π via the identification of Laplace–Beltrami and Casimir operators and Freudenthal's formula for the eigenvalues, which is a particular case of Freudenthal's multiplicity formula, see Humphreys [23], Section 22. Note that eigenvalues corresponding to distinct representations can coincide.

The function $\chi_\pi(\mathbf{A}) = \mathrm{tr}(\pi(\mathbf{A}))$ is called the *character* of the representation corresponding to the highest weight π. Its values can be calculated by means of the Weyl character formula as the ratio of two explicitly known polynomials depending only on the eigenvalues of $\mathbf{A}$. The calculation of these polynomials involves a fair bit of Lie-theoretic machinery, see Bröcker and Tom Dieck [9], Chapter VI, Section 1, or Humphreys [23], Section 24 for details.

Finally, the value $d_\pi = \chi_\pi(\mathbf{I})$ evaluates to the trace of the identity matrix and is thus the dimension of the representation π. Moreover, by the Peter–Weyl Theorem—see Bröcker and Tom Dieck [9], Chapter III—each irreducible representation π (of dimension n) corresponds to an n^2-dimensional eigenspace for λ_π.

Thanks to the integral representation of Matérn kernels given by (29), we may obtain an expression for them without diving into details on how the simplified expression (5) was actually obtained in prior works. Apart from the normalizing constant, the only part of (5) that depends on the length scale, which is the variable with respect to which the integration is performed in (29), is the coefficient $e^{-\frac{\kappa^2}{2}\lambda_\pi}$. Swapping the order of integration and summation, we readily obtain the following formula for Matérn kernels:

$$k_{\nu,\kappa,\sigma^2}(\mathbf{X}, \mathbf{Y}) = \frac{\sigma^2}{C_\nu} \sum_{\pi} \left(\frac{2\nu}{\kappa^2} + \lambda_\pi\right)^{-\nu-\frac{d}{2}} d_\pi \chi_\pi(\mathbf{X}\mathbf{Y}^{-1}). \tag{34}$$

B.5 SPD manifold as $\mathrm{GL}(d)/\mathrm{O}(d)$ and the corresponding heat kernel

First, we prove that the map $T : \mathrm{GL}(d)/\mathrm{O}(d) \to \mathcal{S}^d_{++}$ such that $T(\mathbf{A} \cdot \mathrm{O}(d)) = \mathbf{A}\mathbf{A}^T$ is well defined, i.e. does not depend on the choice of the representative $\mathbf{A}$ of coset $\mathbf{A} \cdot \mathrm{O}(d)$, and is bijective (and in particular, one-to-one). The mapping $\mathrm{GL(d)} \to \mathcal{S}^d_{++}$ acting as $\mathbf{A} \mapsto \mathbf{A}\mathbf{A}^\top$ is surjective because of the Cholesky decomposition. On the other hand, because of the QR decomposition and uniqueness of the Cholesky decomposition, $\mathbf{A}\mathbf{A}^\top = \mathbf{B}\mathbf{B}^\top$ if and only if $\mathbf{A}\mathbf{B}^{-1}$ is an orthogonal matrix. This means that T is correctly defined and is bijective. The identification of $\mathrm{GL}(d)/\mathrm{O}(d)$ and $\mathcal{S}^d_{++}$ through operator T induces [22] the affine-invariant [48] metric over $\mathcal{S}^d_{++}$ with geodesic distance given by

$$\mathrm{dist}(\mathbf{X}, \mathbf{Y}) = \left\| \log(\mathbf{X}^{-1/2}\mathbf{Y}\mathbf{X}^{-1/2}) \right\|, \tag{35}$$

where $\|\mathbf{Z}\| = \sqrt{\mathrm{tr}(\mathbf{Z}\mathbf{Z}^\top)}$ is the Frobenius matrix norm.

For this particular choice of the metric the explicit formulas for the heat kernel in cases $d = 2$ and $d = 3$ are given by Sawyer [43] with the more detailed exposition given in the thesis of Sawyer [44]. There, after observing that the kernel $k(\mathbf{A}, \mathbf{B})$ only depends on the matrix $\mathbf{A}\mathbf{B}^{-1}$, the explicit formulas for the one-parameter function are inferred by a particular technical argument given in that work. In the case $d = 2$, this yields the formula (8) presented in the main text. In the case $d = 3$, an explicit formula is provided as well, although, due to its cumbersome nature, we do not carry it over to this paper.

References

[1] P. A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2007. URL: https://press.princeton.edu/absil. Cited on page 3.

[2] P.-A. Absil, C. G. Baker, and K. A. Gallivan. Trust-Region Methods on Riemannian Manifolds. *Foundations of Computational Mathematics*, 7:303–330, 2007. URL: https://doi.org/10.1007/s10208-005-0179-9. Cited on page 3.

[3] A. Ajoudani, N. G. Tsagarakis, and A. Bicchi. Choosing Poses for Force and Stiffness Control. *IEEE Transactions on Robotics*, 33(6):1483–1490, 2017. URL: https://doi.org/10.1109/TRO.2017.2708087. Cited on page 6.

[4] R. Akrour, D. Sorokin, J. Peters, and G. Neumann. Local Bayesian Optimization of Motor Skills. In *International Conference on Machine Learning*, pages 41–50, 2017. URL: http://proceedings.mlr.press/v70/akrour17a.html. Cited on page 1.

[5] R. Antonova, A. Rai, and C. Atkeson. Deep Kernels for Optimizing Locomotion Controllers. In *Conference on Robot Learning*, pages 47–56, 2017. URL: http://proceedings.mlr.press/v78/antonova17a.html. Cited on page 1.

[6] M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy. BoTorch: Programmable Bayesian Optimization in PyTorch. *arXiv:1910.06403*, 2019. URL: https://botorch.org/. Cited on page 7.

[7] V. Borovitskiy, I. Azangulov, A. Terenin, P. Mostowsky, M. P. Deisenroth, and N. Durrande. Matérn Gaussian Processes on Graphs. In *Artificial Intelligence and Statistics*, pages 2593–2601, 2021. URL: http://proceedings.mlr.press/v130/borovitskiy21a.html. Cited on page 3.

[8] V. Borovitskiy, A. Terenin, P. Mostowsky, and M. Deisenroth. Matérn Gaussian Processes on Riemannian Manifolds. In *Advances in Neural Information Processing Systems*, pages 12426–12437, 2020. URL: https://papers.neurips.cc/paper/2020/hash/92bf5e6240737e0326ea59846a83e076-Abstract.html. Cited on pages 2–6, 14, 15, 18.

[9] T. Bröcker and T. Tom Dieck. Representations of compact Lie groups. In *Graduate Texts in Mathematics*. Springer, 2013. URL: https://doi.org/10.1007/978-3-662-12918-0. Cited on page 18.

[10] P. Chiacchio. Exploiting Redundancy in Minimum-Time Path Following Robot Control. In *American Control Conference*, pages 2313–2318, 1990. URL: https://doi.org/10.23919/ACC.1990.4791142. Cited on page 8.

[11] P. Corke. Robotics Toolbox for Python. URL: https://github.com/petercorke/robotics-toolbox-python. Cited on page 7.

[12] E. De Vito, N. Mücke, and L. Rosasco. Reproducing kernel Hilbert spaces on manifolds: Sobolev and diffusion spaces. *Analysis and Applications*:1–34, 2020. URL: https://doi.org/10.1142/S0219530520400114. Cited on page 15.

[13] A. Feragen and A. Fuster. Geometries and Interpolations for Symmetric Positive Definite Matrices. In *Modeling, Analysis, and Visualization of Anisotropy*, pages 85–113. Springer, 2017. URL: https://doi.org/10.1007/978-3-319-61358-1_5. Cited on page 6.

[14] A. Feragen, F. Lauze, and S. Hauberg. Geodesic exponential kernels: When curvature and linearity conflict. In *Conference on Computer Vision and Pattern Recognition*, pages 3032–3042, 2015. URL: https://doi.org/10.1109/CVPR.2015.7298922. Cited on page 3.

[15] L. P. Fröhlich, M. N. Zeilinger, and E. D. Klenske. Cautious Bayesian Optimization for Efficient and Scalable Policy Search. In *Conference on Learning for Dynamics and Control*, pages 227–240, 2021. URL: http://proceedings.mlr.press/v144/frohlich21a.html. Cited on page 1.

[16] J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson. GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration. In *Advances in Neural Information Processing Systems*, 2018. URL: https://gpytorch.ai/. Cited on page 7.

[17] C. S. Gordon. Survey of Isospectral Manifolds. In *Handbook of Differential Geometry*. Volume 1, pages 747–778. Elsevier, 2000. URL: https://doi.org/10.1016/S1874-5741(00)80009-6. Cited on page 17.

[18] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, 7th edition, 2014. URL: https://doi.org/10.1016/C2010-0-64839-5. Cited on page 15.

[19] A. Grigoryan and M. Noguchi. The heat kernel on hyperbolic space. *Bulletin of the London Mathematical Society*, 30(6):643–650, 1998. URL: https://doi.org/10.1112/S0024609398004780. Cited on page 6.

[20] Y. Gu. Analysis of orientation representations by Lie algebra in robotics. In *International Conference on Robotics and Automation*, pages 874–879, 1988. URL: https://doi.org/10.1109/ROBOT.1988.12170. Cited on page 2.

[21] L. Guilamo, J. Kuffner, K. Nishiwaki, and S. Kagami. Manipulability optimization for trajectory generation. In *International Conference on Robotics and Automation*, pages 2017–2022, 2006. URL: https://doi.org/10.1109/ROBOT.2006.1642001. Cited on page 8.

[22] C. S. Herz. Les espaces symétriques pour piétons. *Publications mathématiques d'Orsay, Séminaire d'analyse harmonique 79-07*, 1979. URL: https://www.imo.universite-paris-saclay.fr/~biblio/numerisation/docs/31_SEMINAIRE/pdf/31_SEMINAIRE.pdf. Cited on page 19.

[23] J. Humphreys. *Introduction to Lie Algebras and Representation Theory*, volume 9. Springer, 1994. URL: https://doi.org/10.1007/978-1-4612-6398-2. Cited on page 18.

[24] V. C. Ines Chami Albert Gu and C. Ré. From Trees to Continuous Embeddings and Back: Hyperbolic Hierarchical Clustering. In *Advances in Neural Information Processing Systems*, pages 15065–15076, 2020. URL: https://papers.neurips.cc/paper/2020/hash/ac10ec1ace51b2d973cd87973a98d3ab-Abstract.html. Cited on page 6.

[25] N. Jaquier and L. Rozo. High-Dimensional Bayesian Optimization via Nested Riemannian Manifolds. In *Advances in Neural Information Processing Systems*, pages 20939–20951, 2020. URL: https://papers.neurips.cc/paper/2020/hash/f05da679342107f92111ad9d65959cd3-Abstract.html. Cited on page 3.

[26] N. Jaquier, L. Rozo, D. G. Caldwell, and S. Calinon. Geometry-aware Manipulability Learning, Tracking and Transfer. *International Journal of Robotics Research*, 20(2-3):624–650, 2021. URL: https://doi.org/10.1177/0278364920946815. Cited on pages 8, 13.

[27] N. Jaquier, L. Rozo, S. Calinon, and M. Bürger. Bayesian Optimization meets Riemannian Manifolds in Robot Learning. In *Conference on Robot Learning*, pages 233–246, 2019. URL: http://proceedings.mlr.press/v100/jaquier20a.html. Cited on pages 2–4, 7.

[28] L. E. Kavraki and S. M. LaValle. Motion Planning. In B. Siciliano and O. Khatib, editors, *Springer Handbook of Robotics*, pages 139–162. 2016. URL: https://doi.org/10.1007/978-3-319-32552-1_7. Cited on page 2.

[29] I. Lee and J.-H. Oh. Humanoid Posture Selection for Reaching Motion and a Cooperative Balancing Controller. *Journal of Intelligent and Robotic Systems*, 81:301–316, 2016. URL: https://doi.org/10.1007/s10846-015-0225-z. Cited on page 8.

[30] F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011. URL: https://doi.org/10.1111/j.1467-9868.2011.00777.x. Cited on pages 2, 3.

[31] G. J. Lord, C. E. Powell, and T. Shardlow. *An introduction to computational stochastic PDEs*. Cambridge University Press, 2014. URL: https://doi.org/10.1017/CBO9781139017329. Cited on page 4.

[32] R. Marchant and F. Ramos. Bayesian Optimisation for informative continuous path planning. In *International Conference on Robotics and Automation*, pages 6136–6143, 2014. URL: https://doi.org/10.1109/ICRA.2014.6907763. Cited on page 8.

[33] A. Marco, P. Hennig, J. Bohg, and S. S. and S. Trimpe. Automatic LQR Tuning Based on Gaussian Process Global Optimization. In *International Conference on Robotics and Automation*, pages 270–277, 2016. URL: https://doi.org/10.1109/ICRA.2016.7487144. Cited on pages 1, 3, 6.

[34] A. Marco, P. Hennig, S. Schaal, and S. Trimpe. On the Design of LQR Kernels for Efficient Controller Learning. In *Conference on Decision and Control*, pages 5193–5200, 2017. URL: https://doi.org/10.1109/CDC.2017.8264429. Cited on page 1.

[35] J. Močkus. On Bayesian methods for seeking the extremum. In *Optimization Techniques IFIP Technical Conference*, pages 400–404, 1975. URL: https://doi.org/10.1007/3-540-07165-2_55. Cited on page 7.

[36] M. Nickel and D. Kiela. Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry. In *International Conference on Machine Learning*, pages 3779–3788, 2018. URL: http://proceedings.mlr.press/v80/nickel18a.html. Cited on page 8.

[37] M. Nickel and D. Kiela. Poincaré Embeddings for Learning Hierarchical Representations. In *Advances in Neural Information Processing Systems*, 2017. URL: https://papers.neurips.cc/paper/2017/hash/59dfa2df42d9e3d41f5b02bfc32229dd-Abstract.html. Cited on page 6.

[38] A. Rai, R. Antonova, S. Song, W. Martin, H. Geyer, and C. Atkeson. Bayesian Optimization Using Domain Knowledge on the ATRIAS Biped. In *International Conference on Robotics and Automation*, pages 1771–1778, 2018. URL: https://doi.org/10.1109/ICRA.2018.8461237. Cited on page 1.

[39] C. E. Rasmussen and C. K. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. URL: http://www.gaussianprocess.org/gpml/. Cited on page 15.

[40] L. Roveda, M. Magni, M. Cantoni, D. Piga, and G. Bucca. Human–robot collaboration in sensorless assembly task learning enhanced by uncertainties adaptation via Bayesian Optimization. *Robotics and Autonomous Systems*, 136:103711, 2021. URL: https://doi.org/10.1016/j.robot.2020.103711. Cited on page 3.

[41] L. Rozo. Interactive Trajectory Adaptation through Force-guided Bayesian Optimization. In *International Conference on Intelligent Robots and Systems*, pages 7596–7603, 2019. URL: https://doi.org/10.1109/IROS40897.2019.8968571. Cited on page 3.

[42] K. A. Saar, F. Giardina, and F. Iida. Model-Free Design Optimization of a Hopping Robot and Its Comparison With a Human Designer. *IEEE Robotics and Automation Letters*, 3(2):1245–1251, 2018. URL: https://doi.org/10.1109/LRA.2018.2795646. Cited on page 1.

[43] P. Sawyer. The heat equation on the spaces of positive definite matrices. *Canadian Journal of Mathematics*, 44(3):624–651, 1992. URL: https://doi.org/10.4153/CJM-1992-038-7. Cited on pages 6, 19.

[44] P. Sawyer. *The Heat Equation on the Symmetric Space Associated with* $\mathrm{SL}(n, \mathbb{R})$. PhD thesis, McGill University Libraries, 1989. URL: https://central.bac-lac.gc.ca/.item?id=TC-QMM-74269&op=pdf&app=Library&oclc_number=897817061. Cited on page 19.

[45] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016. URL: https://doi.org/10.1109/JPROC.2015.2494218. Cited on pages 2, 3.

[46] J. Snoek, H. Larochelle, and R. P. Adams. Practical Bayesian Optimization of Machine Learning Algorithms. In *Advances in Neural Information Processing Systems*, 2012. URL: https://papers.neurips.cc/paper/2012/hash/05311655a15b75fab86956663e1819cd-Abstract.html. Cited on page 1.

[47] E. M. Stein. *Topics in Harmonic Analysis, Related to the Littlewood-Paley Theory*. Princeton University Press, 1970. URL: https://doi.org/10.1515/9781400881871. Cited on page 18.

[48] Y. Thanwerdas and X. Pennec. Is affine-invariance well defined on SPD matrices? A principled continuum of metrics. In *International Conference on Geometric Science of Information*, pages 502–510. Springer, 2019. URL: https://doi.org/10.1007/978-3-030-26980-7_52. Cited on page 19.

[49] J. Townsend, N. Koep, and S. Weichwald. Pymanopt: A Python Toolbox for Optimization on Manifolds using Automatic Differentiation. *Journal of Machine Learning Research*, 17(137):1–5, 2016. URL: http://jmlr.org/papers/v17/16-177.html. Cited on page 7.

[50] F. Tronarp, T. Karvonen, and S. Särkkä. Mixture Representation of the Matérn Class with Applications in State Space Approximations and Bayesian Quadrature. In *International Workshop on Machine Learning for Signal Processing*, pages 1–6, 2018. URL: https://doi.org/10.1109/MLSP.2018.8516992. Cited on page 3.

[51] J.-Y. Wen and K. Kreutz-Delgado. The attitude control problem. *IEEE Transactions on Automatic Control*, 36(10):1148–1162, 1991. URL: https://doi.org/10.1109/9.90228. Cited on page 2.

[52] P. Whittle. Stochastic processes in several dimensions. *Bulletin of the International Statistical Institute*, 40(2):974–994, 1963. Cited on page 4.

[53] A. Wilson, A. Fern, and P. Tadepalli. Using Trajectory Data to Improve Bayesian Optimization for Reinforcement Learning. *Journal of Machine Learning Research*, 15(8):253–282, 2014. URL: http://jmlr.org/papers/v15/wilson14a.html. Cited on page 1.

[54] J. T. Wilson, V. Borovitskiy, A. Terenin, P. Mostowsky, and M. P. Deisenroth. Efficiently Sampling Functions from Gaussian Process Posteriors. In *International Conference on Machine Learning*, pages 10292–10302, 2020. URL: http://proceedings.mlr.press/v119/wilson20a.html. Cited on page 4.

[55] J. T. Wilson, V. Borovitskiy, A. Terenin, P. Mostowsky, and M. P. Deisenroth. Pathwise Conditioning of Gaussian Processes. *Journal of Machine Learning Research*, 22(105):1–47, 2021. URL: https://www.jmlr.org/papers/v22/20-1260.html. Cited on page 4.

[56] M. Wong. Weyl Transforms, Heat Kernels, Green Functions and Riemann Zeta Functions on Compact Lie Groups. In *Modern Trends in Pseudo-Differential Operators*, pages 67–85. Springer, 2006. URL: https://doi.org/10.1007/978-3-7643-8116-5_4. Cited on page 18.

[57] T. Yoshikawa. Manipulability of Robotic Mechanisms. *The International Journal of Robotics Research*, 4(2):3–9, 1985. URL: https://doi.org/10.1177/027836498500400201. Cited on page 6.