A. Proofs

Proposition 1. *Let k be a differentiable positive definite kernel. Then* $\nabla_x k(x,x)|_{x=y} - 2\nabla_x k(x,y)|_{x=y} = 0$.

Proof. Since k is a positive definite kernel there exists a Hilbert space $\mathcal{H}$ and a feature map $\varphi : \mathbb{R}^d \to \mathcal{H}$, such that $k(x,x') = \langle \varphi(x), \varphi(x') \rangle_{\mathcal{H}}$. Consider first the map $\tau : \mathbb{R}^d \to \mathbb{R}$, defined by $\tau(x) = k(x,x)$. We write $\tau = \psi \circ \varphi$, where $\psi : \mathcal{H} \to \mathbb{R}$, $\psi(f) = \|f\|^2_{\mathcal{H}}$. We can obtain $\nabla_x k(x,x)|_{x=y}$ from the Fréchet derivative $D\tau(y) \in \mathcal{B}(\mathbb{R}^d, \mathbb{R})$ of τ at y, which to each $y \in \mathbb{R}^d$ associates a bounded linear operator from $\mathbb{R}^d$ to $\mathbb{R}$ (Steinwart & Christmann, 2008, Definition A.5.14). By the chain rule for

Fréchet derivatives (Steinwart & Christmann, 2008, Lemma A.5.15(b)), the value of $D\tau(y)$ at some $x' \in \mathbb{R}^d$ is

$$[D\tau(y)](x') = [D\psi(\varphi(y)) \circ D\varphi(y)](x'),$$

where $D\varphi(y) \in \mathcal{B}(\mathbb{R}^d, \mathcal{H})$, and $D\psi(\varphi(y)) \in \mathcal{B}(\mathcal{H}, \mathbb{R})$. The derivative $D\varphi$ of the feature map exists whenever k is a differentiable function (Steinwart & Christmann, 2008, Section 4.3). It is readily shown that $D\psi[\varphi(y)] = 2\langle \varphi(y), \cdot \rangle_{\mathcal{H}}$, so that

$$[D\tau(y)](x') = 2\langle \varphi(y), [D\varphi(y)](x') \rangle_{\mathcal{H}}.$$

Next, we consider the map $\kappa_y(x) = k(x,y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}$, i.e., $\kappa_y = \psi_y \circ \varphi$ where $\psi_y(f) = \langle f, \varphi(y) \rangle_{\mathcal{H}}$. Since ψ_y is a linear scalar function on $\mathcal{H}$, $D\psi_y(f) = \langle \varphi(y), \cdot \rangle_{\mathcal{H}}$. Again, by the chain rule:

$$\begin{aligned} [D\kappa_y(y)](x') &= [D\psi_y(\varphi(y)) \circ D\varphi(y)](x') \\ &= \langle \varphi(y), [D\varphi(y)](x') \rangle_{\mathcal{H}}, \end{aligned}$$

and thus $(D\tau(y) - 2D\kappa_y(y))(x') = 0$, for all $x' \in \mathbb{R}^d$, and we obtain equality of operators. Since Fréchet derivatives can also be written as inner products with the gradients, $(\nabla_x k(x,x)|_{x=y} - 2\nabla_x k(x,y)|_{x=y})^\top x' = (D\tau(y) - 2D\kappa_y(y))(x') = 0, \forall x' \in \mathbb{R}^d$, which proves the claim. □

Proposition 2. $q_{\mathbf{z}}(\cdot|y) = \mathcal{N}(y, \gamma^2 I + \nu^2 M_{\mathbf{z},y} H M_{\mathbf{z},y}^\top)$.

Proof. We start with

$$\begin{aligned} p(\beta)p(x^*|y,\beta) &= \frac{1}{(2\pi)^{\frac{n+d}{2}} \gamma^d \nu^n} \exp\left(-\frac{1}{2\nu^2}\beta^\top\beta\right) \\ &\quad \cdot \exp\left(-\frac{1}{2\gamma^2}(x^* - y - M_{\mathbf{z},y}H\beta)^\top (x^* - y - M_{\mathbf{z},y}H\beta)\right) \\ &= \frac{1}{(2\pi)^{\frac{n+d}{2}} \gamma^d \nu^n} \exp\left(-\frac{1}{2\gamma^2}(x^* - y)^\top (x^* - y)\right) \\ &\quad \cdot \exp\left(-\frac{1}{2}\left(\beta^\top\left(\frac{1}{\nu^2}I + \frac{1}{\gamma^2}HM_{\mathbf{z},y}^\top M_{\mathbf{z},y}H\right)\beta - \frac{2}{\gamma^2}\beta^\top HM_{\mathbf{z},y}^\top (x^* - y)\right)\right). \end{aligned}$$

Now, we set

$$\begin{aligned} \Sigma^{-1} &= \frac{1}{\nu^2}I + \frac{1}{\gamma^2}HM_{\mathbf{z},y}^\top M_{\mathbf{z},y}H \\ \mu &= \frac{1}{\gamma^2}\Sigma HM_{\mathbf{z},y}^\top (x^* - y), \end{aligned}$$

and application of the standard Gaussian integral

$$\int \exp\left(-\frac{1}{2}\left(\beta^\top\Sigma^{-1}\beta - 2\beta^\top\Sigma^{-1}\mu\right)\right) d\beta = (2\pi)^{n/2}\sqrt{\det\Sigma}\exp\left(\frac{1}{2}\mu^\top\Sigma^{-1}\mu\right),$$

leads to

$$q_{\mathbf{z}}(x^*|y) = \frac{\sqrt{\det \Sigma}}{(2\pi)^{\frac{d}{2}} \gamma^d \nu^n} \exp\left(-\frac{1}{2\gamma^2}(x^*-y)^\top (x^*-y)\right) \cdot \exp\left(\frac{1}{2}\mu^\top \Sigma^{-1}\mu\right).$$

This is just a d-dimensional Gaussian density where both the mean and covariance will, in general, depend on y. Let us consider the exponent

$$\begin{aligned}
&-\frac{1}{2\gamma^2}(x^*-y)^\top (x^*-y) + \frac{1}{2}\mu^\top \Sigma^{-1}\mu = \\
&-\frac{1}{2}\Bigg\{\frac{1}{\gamma^2}(x^*-y)^\top (x^*-y) \\
&\qquad -\frac{1}{\gamma^4}(x^*-y)^\top M_{\mathbf{z},y} H\Sigma H M_{\mathbf{z},y}^\top (x^*-y)\Bigg\} = \\
&-\frac{1}{2}\left\{(x^*-y)^\top R^{-1}(x^*-y)\right\},
\end{aligned}$$

where $R^{-1} = \frac{1}{\gamma^2}(I - \frac{1}{\gamma^2} M_{\mathbf{z},y} H\Sigma H M_{\mathbf{z},y}^\top)$. We can simplify the covariance R using the Woodbury identity to obtain:

$$\begin{aligned}
R &= \gamma^2 (I - \frac{1}{\gamma^2} M_{\mathbf{z},y} H\Sigma H M_{\mathbf{z},y}^\top)^{-1} \\
&= \gamma^2 \left(I + M_{\mathbf{z},y} H \left(\gamma^2 \Sigma^{-1} - H M_{\mathbf{z},y}^\top M_{\mathbf{z},y} H\right)^{-1} H M_{\mathbf{z},y}^\top\right) \\
&= \gamma^2 \left(I + M_{\mathbf{z},y} H \left(\frac{\gamma^2}{\nu^2} I\right)^{-1} H M_{\mathbf{z},y}^\top\right) \\
&= \gamma^2 I + \nu^2 M_{\mathbf{z},y} H M_{\mathbf{z},y}^\top.
\end{aligned}$$

Therefore, the proposal density is $q_{\mathbf{z}}(\cdot|y) = \mathcal{N}(y, \gamma^2 I + \nu^2 M_{\mathbf{z},y} H M_{\mathbf{z},y}^\top)$. □

B. Further details on synthetic experiments

Proposal contours for the Flower target. The d-dimensional flower target $\mathcal{F}(r_0, A, \omega, \sigma)$ is given by

$$\mathcal{F}(x; r_0, A, \omega, \sigma) = \exp\left(-\frac{\sqrt{x_1^2 + x_2^2} - r_0 - A\cos(\omega \text{atan2}(x_2, x_1))}{2\sigma^2}\right) \mathcal{N}(x_{3:d}; 0, I).$$

This distribution concentrates around the r_0-circle with a periodic perturbation (with amplitude A and frequency ω) in the first two dimensions. For $A = 0$, we obtain a band around the r_0-circle, which we term the ring target. Figure 6 gives the contour plots of the MCMC Kameleon proposal distributions on two instances of the flower target.

Convergence statistics for the Banana target. Figure 7 illustrates how the norm of the mean and quantile deviation (shown for 0.5-quantile) for the strongly twisted Banana target decrease as a function of the number of iterations. This shows that the trends observed in the main text persist along the evolution of the whole chain.

C. Principal Components Proposals

An alternative approach to the standard adaptive Metropolis, discussed in Andrieu & Thoms (2008, Algorithm 8), is to extract $m \leq d$ principal eigenvalue-eigenvector pairs $\{(\lambda_j, v_j)\}_{j=1}^m$ from the estimated covariance matrix $\Sigma_{\mathbf{z}}$ and use the proposal that takes form of a mixture of one-dimensional random walks along the principal eigendirections

$$q_{\mathbf{z}}(\cdot|y) = \sum_{j=1}^m \omega_j \mathcal{N}(y, \nu_j^2 \lambda_j v_j v_j^\top). \tag{7}$$

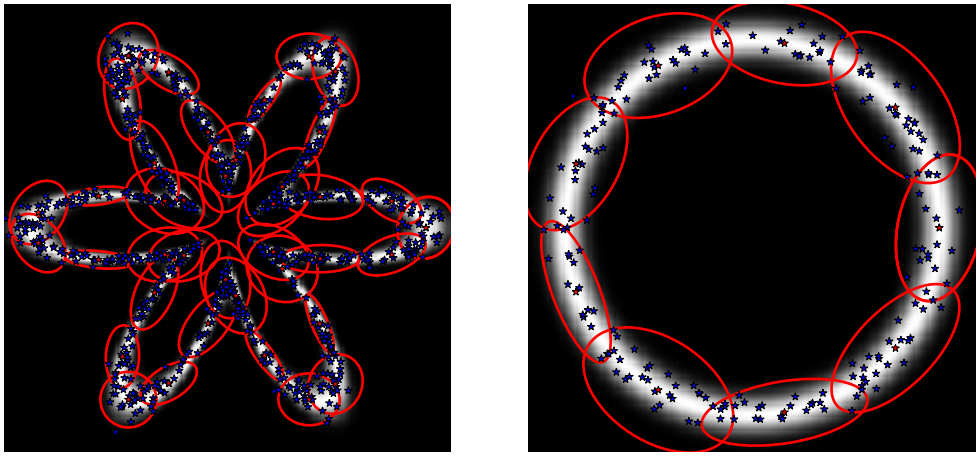

Figure 6. 95% contours (red) of proposal distributions evaluated at a number of points, for the flower and the ring target. Underneath are the density heatmaps, and the samples (blue) used to construct the proposals.

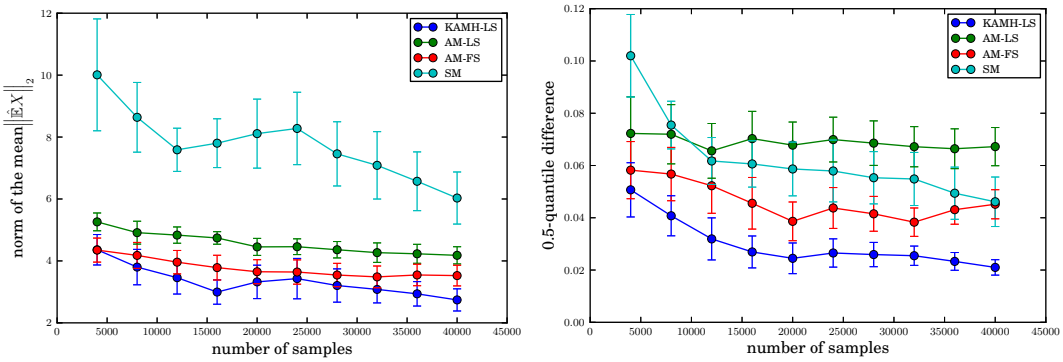

Figure 7. Comparison of **SM**, **AM-FS**, **AM-LS** and **KAMH-LS** in terms of the norm of the estimated mean (left) and in terms of the deviation from the 0.5-quantile (right) on the strongly twisted Banana distribution. The results are averaged over 20 chains for each sampler. Error bars represent 80%-confidence intervals.

In other words, given the current chain state y, the j-th principal eigendirection is chosen with probability ω_j (choice $\omega_j = \lambda_j / \sum_{l=1}^m \lambda_l$ is suggested), and the proposed point is

$$x^* = y + \rho \nu_j \sqrt{\lambda_j} v_j, \tag{8}$$

with $\rho \sim \mathcal{N}(0,1)$. Note that each eigendirection may have a different scaling factor ν_j in addition to the scaling with the eigenvalue.

We can consider an analogous version of the update (8) performed in the RKHS

$$f \quad = \quad k(\cdot, y) + \rho \nu_j \sqrt{\lambda_j} \mathbf{v}_j, \tag{9}$$

with $m \leq n$ principal eigenvalue-eigenfunction pairs $\{(\lambda_j, \mathbf{v}_j)\}_{j=1}^m$. It is readily shown that the eigenfunctions $\mathbf{v}_j = \sum_{i=1}^n \tilde{\alpha}_i^{(j)} \left[k(\cdot, z_i) - \mu_{\mathbf{z}}\right]$ lie in the subspace $\mathcal{H}_{\mathbf{z}}$ induced by $\mathbf{z}$, and that the coefficients vectors $\tilde{\alpha}^{(j)} = \left(\tilde{\alpha}_1^{(j)} \cdots \tilde{\alpha}_n^{(j)}\right)^\top$ are proportional to the eigenvector of the centered kernel matrix HKH, with normalization chosen so that $\|\mathbf{v}_j\|_{\mathcal{H}_k}^2 = \left(\tilde{\alpha}^{(j)}\right)^\top HKH\tilde{\alpha}^{(j)} = \lambda_j \left\|\tilde{\alpha}^{(j)}\right\|_2^2 = 1$ (so that the eigenfunctions have the unit RKHS norm). Therefore, the update (9) has form

$$f = k(\cdot, y) + \sum_{i=1}^n \beta_i^{(j)} \left[k(\cdot, z_i) - \mu_{\mathbf{z}}\right],$$

where $\beta^{(j)} = \rho \nu_j \sqrt{\lambda_j} \tilde{\alpha}^{(j)}$. But $\alpha^{(j)} = \sqrt{\lambda_j} \tilde{\alpha}^{(j)}$ are themselves the (unit norm) eigenvectors of HKH, as $\left\|\alpha^{(j)}\right\|_2^2 = \lambda_j \left\|\tilde{\alpha}^{(j)}\right\|_2^2 = 1$. Therefore, the appropriate scaling with eigenvalues is already included in the β-coefficients, just like in the MCMC Kameleon, where the β-coefficients are isotropic.

Now, we can construct the MCMC PCA-Kameleon by simply substituting β-coefficients with $\rho \nu_j \alpha^{(j)}$, where j is the selected eigendirection, and ν_j is the scaling factor associated to the j-th eigendirection. We have the following steps:

1. Perform eigendecomposition of HKH to obtain the $m \leq n$ eigenvectors $\{\alpha_j\}_{j=1}^m$.
2. Draw $j \sim \text{Discrete}\left[\omega_1, \ldots, \omega_m\right]$
3. $\rho \sim \mathcal{N}(0,1)$
4. $x^* | y, \rho, j \sim \mathcal{N}(y + \rho \nu_j M_{\mathbf{z},y} H \alpha^{(j)}, \gamma^2 I)$ ($d \times 1$ normal in the original space)

Similarly as before, we can simplify the proposal by integrating out the scale ρ of the moves in the RKHS.

Proposition 3. $q_{\mathbf{z}}(\cdot|y) = \sum_{j=1}^m \omega_j \mathcal{N}(y, \gamma^2 I + \nu_j^2 M_{\mathbf{z},y} H \alpha^{(j)} \left(\alpha^{(j)}\right)^\top H M_{\mathbf{z},y}^\top)$.

Proof. We start with

$$\begin{aligned} p(\rho) p(x^* \mid y, \rho, j) \quad \propto \quad & \exp\left[-\frac{1}{2\gamma^2}(x^* - y)^\top (x^* - y)\right] \\ & \cdot \exp\left[-\frac{1}{2}\left\{\left(1 + \frac{\nu_j^2}{\gamma^2}\left(\alpha^{(j)}\right)^\top H M_{\mathbf{z},y}^\top M_{\mathbf{z},y} H \alpha^{(j)}\right)\rho^2 - 2\rho \frac{\nu_j}{\gamma^2}\left(\alpha^{(j)}\right)^\top H M_{\mathbf{z},y}^\top (x^* - y)\right\}\right]. \end{aligned}$$

By substituting

$$\begin{aligned} \sigma^{-2} \quad &= \quad 1 + \frac{\nu_j^2}{\gamma^2}\left(\alpha^{(j)}\right)^\top H M_{\mathbf{z},y}^\top M_{\mathbf{z},y} H \alpha^{(j)}, \\ \mu \quad &= \quad \sigma^2 \left(\frac{\nu_j}{\gamma^2}\left(\alpha^{(j)}\right)^\top H M_{\mathbf{z},y}^\top (x^* - y)\right), \end{aligned}$$

we integrate out ρ to obtain:

$$\begin{aligned} p\left(x^* \mid y, j\right) &\propto \exp\left[-\frac{1}{2}\left\{\frac{1}{\gamma^2}(x^*-y)^\top(x^*-y) - \frac{\nu_j^2\sigma^2}{\gamma^4}\left(x^*-y\right)^\top M_{\mathbf{z},y} H\alpha^{(j)}\left(\alpha^{(j)}\right)^\top H M_{\mathbf{z},y}^\top\left(x^*-y\right)\right\}\right] \\ &= \exp\left[-\frac{1}{2}\left(x^*-y\right)^\top R^{-1}\left(x^*-y\right)\right] \end{aligned}$$

where $R^{-1} = \frac{1}{\gamma^2}\left(I - \frac{\nu_j^2\sigma^2}{\gamma^2} M_{\mathbf{z},y} H\alpha^{(j)}\left(\alpha^{(j)}\right)^\top H M_{\mathbf{z},y}^\top\right)$. We can simplify the covariance R using the Woodbury identity to obtain:

$$\begin{aligned} R &= \gamma^2(I - \frac{\nu_j^2\sigma^2}{\gamma^2} M_{\mathbf{z},y} H\alpha^{(j)}\left(\alpha^{(j)}\right)^\top H M_{\mathbf{z},y}^\top)^{-1} \\ &= \gamma^2\left(I + \frac{\nu_j^2\sigma^2}{\gamma^2} M_{\mathbf{z},y} H\alpha^{(j)}\left(1 - \frac{\nu_j^2\sigma^2}{\gamma^2}\left(\alpha^{(j)}\right)^\top H M_{\mathbf{z},y}^\top M_{\mathbf{z},y} H\alpha^{(j)}\right)^{-1}\left(\alpha^{(j)}\right)^\top H M_{\mathbf{z},y}^\top\right) \\ &= \gamma^2\left(I + \frac{\nu_j^2}{\gamma^2} M_{\mathbf{z},y} H\alpha^{(j)}\left(\alpha^{(j)}\right)^\top H M_{\mathbf{z},y}^\top\right) \\ &= \gamma^2 I + \nu_j^2 M_{\mathbf{z},y} H\alpha^{(j)}\left(\alpha^{(j)}\right)^\top H M_{\mathbf{z},y}^\top. \end{aligned}$$

The claim follows after summing over the choice j of the eigendirection (w.p. ω_j). □