# Kernel Herding

## 1. Finite-dimensional Kernel Herding

### 1.1 Setting.

Let $\mathcal{X} = \{-1, 1\}^d$ and

$$\mathcal{H} := \left\{ f : \mathcal{X} \to \mathbb{R} \mid f(x) = \langle f, \Phi(x) \rangle_{\mathcal{H}}, \text{ and } \Phi(x) = (x, xx^T) \right\}.$$

The feature map $\Phi(x) = (x, xx^T)$ is composed of $x$ and of all of its pairwise products $xx^T$, $x \in \mathcal{X}$.

For $f, g \in \mathcal{H}$,

$$\langle f, g \rangle_{\mathcal{H}} := \sum_{i=1}^{d} f_i(x) f_i(x)$$

defines inner product. Thus, $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ is a Hilbert space. Moreover, the Hilbert space $\mathcal{H}$ is also a RKHS with reproducing kernel $k(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}$ for $x, y \in \mathcal{X}$. Indeed,

$$k(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}} = \sum_{i=1}^{d} \Phi_i(x) \Phi_i(y) = \langle x, y \rangle_2. \tag{1.1}$$

And for $t \in \mathcal{X}$,

$$\Phi(t)(x) = k(t, x) \in \mathcal{H}.$$

is a feature map $\Phi : \mathcal{X} \to \mathcal{H}$ for the kernel $k$ such that

$$k(x, y) = \big\langle k(z, x), k(z, y) \big\rangle_{\mathcal{H}}$$

for $x, y, z \in \mathcal{X}$.

We denote $\mathcal{M} \subset \mathcal{H}$ the marginal polytope as $\mathcal{M} = \text{conv}(\{\Phi(x) \mid x \in \mathcal{X}\}$. In this setting, we compute the expectation

$$\mu(t) := \mathbb{E}_{p(x)} \Phi(x)(t) = \sum_{i=1}^{2^d} p(x_i) \Phi(x_i)(t) \in \mathcal{M}.$$

such that $\sum_{i=1}^{2^d} p(y_i) = 1$ where $x_i \in \{-1, 1\}^d$. Then

$$\begin{aligned}
\langle \mu, \mu \rangle_{\mathcal{H}} &= \Big\langle \sum_{i=1}^{2^d} p(x_i) k(t, x_i), \sum_{j=1}^{2^d} p(x_j) k(t, x_j) \Big\rangle_{\mathcal{H}} \\
&= \sum_{i=1}^{2^d} \sum_{j=1}^{2^d} p(x_i) p(x_j) \big\langle k(t, x_i), k(t, x_j) \big\rangle_{\mathcal{H}} \\
&= \sum_{i=1}^{2^d} \sum_{j=1}^{2^d} p(x_i) p(x_j) k(x_i, x_j) \\
&= \sum_{i=1}^{2^d} \sum_{j=1}^{2^d} p(x_i) p(x_j) \langle x_i, x_j \rangle_2.
\end{aligned}$$

Let $f(x) = \sum_n w_n k(t_n, x)$ and $g(x) = \sum_m w_m k(t_m, x)$, where $t_n, t_m \in \mathcal{X}$ are distinct points such that $k(t_n, x)$ corresponds to a vertex of $\mathcal{M}$. Then,

$$
\begin{aligned}
\langle f, g \rangle_{\mathcal{H}} &= \sum_n \sum_m w_n w_m k(t_n, t_m) \\
&= \sum_n \sum_m w_n w_m \langle t_n, t_m \rangle_2 .
\end{aligned}
$$

Finally, we have

$$
\begin{aligned}
\langle f, \mu \rangle_{\mathcal{H}} &= \Big\langle \sum_n w_n k(t_n, x), \sum_{j=1}^{2^d} p(x_j) k(t, x_j) \Big\rangle_{\mathcal{H}} \\
&= \sum_n \sum_{i=1}^{2^d} w_n p(x_i) k(t_n, x_i) \\
&= \sum_n \sum_{i=1}^{2^d} w_n p(x_i) \langle t_n, x_i \rangle_2 .
\end{aligned}
$$

## 1.2 Experimental Results.

The convergence rate for finite Kernel with any distribution $\mu \in \text{interior}(\mathcal{M})$ is $\frac{1}{T^2}$ for open loop step-size $\gamma_t = \frac{2}{t+2}$ and FW with short step converges linearly.

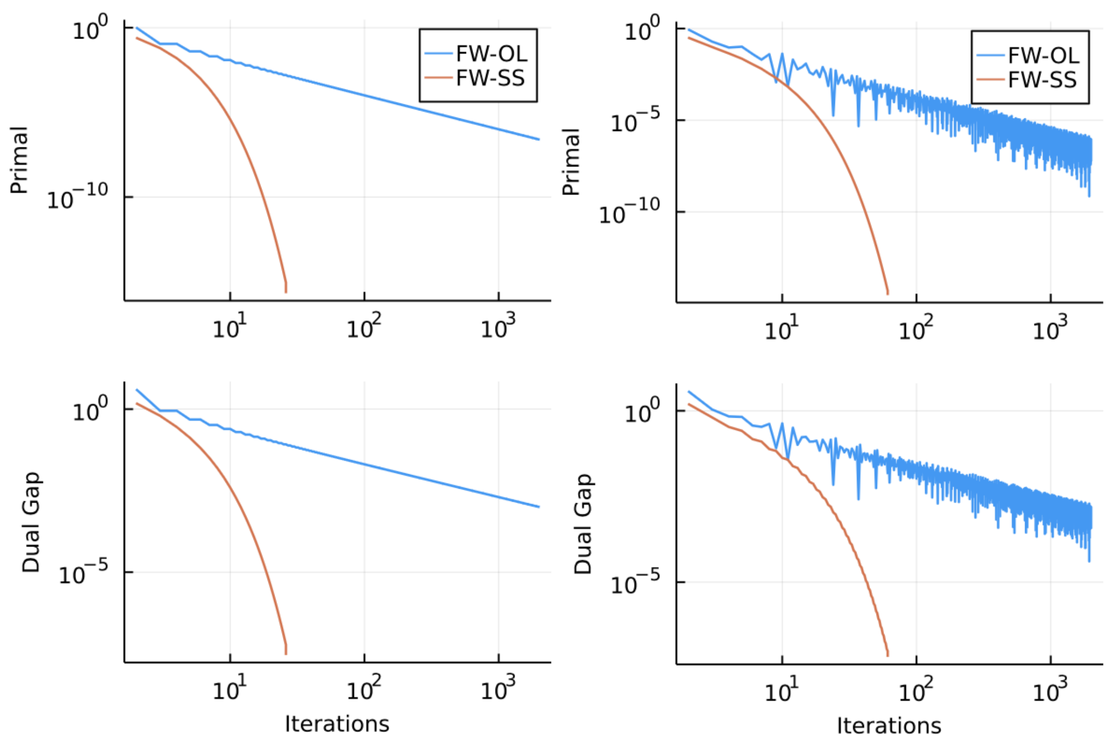


Figure 1: Solving (OPT-KH) with FW with short step (FW-SS) and open loop step-size rules of the form $\gamma_t = \frac{2}{t+2}$ (FW-OL) with uniform distribution (left) and non-uniform distribution (right) for dimension 2.

## 2. Infinite-dimensional Kernel Herding

### 2.1 Matern's Kernel

## References